

**PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR
FACULTAD DE INGENIERÍA
ESCUELA DE SISTEMAS**



**DISERTACIÓN PREVIA A LA OBTENCIÓN DEL TÍTULO DE INGENIERO EN
SISTEMAS Y COMPUTACIÓN**

**“ANÁLISIS COMPARATIVO DE HERRAMIENTAS DE SOFTWARE LIBRE Y
PROPIETARIO PARA LA GESTIÓN DE BIG DATA EN EMPRESAS DE
COMERCIALIZACIÓN MASIVA”**

**AUTOR:
JOHN STEVEN CABEZAS JÁCOME**

**DIRECTOR:
ING. OSWALDO ESPINOSA**

QUITO, NOVIEMBRE 2015

RESUMEN

El concepto de Big Data ha venido tomando un gran impacto en el campo de la tecnología, especialmente en el análisis y gestión de grandes volúmenes de información. Es por esta razón que las empresas hoy en día optan por introducir tal concepto como una oportunidad para explotar al máximo toda esa información que ayudará a mejorar la cultura de toma de decisiones dentro de sus diferentes áreas operativas así como también el manejo de dicha información en cada una de ellas. Como consecuencia es necesario analizar varios conceptos sobre Big Data, entre los que se incluyen definiciones, características, ventajas y desventajas con el fin de tener más claro el panorama de funcionamiento dentro de una organización en particular.

Las empresas de comercialización masiva tienen un fuerte impacto en la economía con la venta de productos y servicios. Como caso particular de estudio en el presente trabajo de disertación de grado es de gran importancia analizar definiciones, características y sobre todo el modo de operación de dichas empresas en el mercado.

La inteligencia de negocios junto con la generación de conocimiento son conceptos fundamentales para el proceso de toma de decisiones y de transformación de la información. Es por este motivo que se analizarán definiciones, objetivos, características y funciones principales de la inteligencia de negocios que sirvan de guía para explicar de una manera adecuada su influencia en la toma de decisiones y en el manejo de información.

Las herramientas de software libre y propietario ofrecen diferentes enfoques de aplicación de acuerdo a la realidad o cultura que se tiene en las empresas. Este trabajo lo que muestra es un análisis comparativo de una herramienta de software libre contra una herramienta de software propietario para la gestión de Big Data en empresas de comercialización masiva con el fin de determinar cuál herramienta es la que mejor se ajusta al concepto o a la realidad de dichas empresas.

DEDICATORIA

El presente trabajo de disertación de grado va dedicado para mis padres, por ser mi soporte y apoyo incondicional en absolutamente todas las actividades que he realizado hasta hoy, puesto que me han dado un ejemplo de perseverancia, lucha y esfuerzo día a día. Sin duda alguna ellos han sido y serán siempre mi mayor fuente de inspiración para conseguir logros tan importantes como éste.

Además, se lo dedico al todopoderoso, por darme todo lo que tengo, por ser la persona que soy y por permitirme estar donde estoy ahora, ya que todo esto se lo debo a él.

No quisiera dejar de lado a los demás miembros de mi familia, que les debo tanto y les agradezco por cada uno de los consejos que me han dado para llegar hasta acá.

AGRADECIMIENTOS

Agradezco a todos mis profesores miembros del tribunal por haber aceptado de manera muy gentil y amable colaborar con mi proyecto de disertación y por todo lo que me han enseñado, tanto en el ámbito académico como en el ámbito personal.

También agradezco a mis amigos que con su apoyo también han sido parte de todo este proceso.

Finalmente agradezco a la PUCE, por brindarme su acogida y por todo lo que he aprendido a lo largo de estos años.

TABLA DE CONTENIDOS

1. CAPÍTULO 1: FORMULACIÓN DEL PROBLEMA.....	- 1 -
1.1. Antecedentes.....	- 1 -
1.2. Justificación.....	- 2 -
1.3. Objetivos.....	- 3 -
1.3.1. <i>Objetivo General</i>	- 3 -
1.3.2. <i>Objetivos Específicos</i>	- 3 -
1.4. Alcance.....	- 3 -
2. CAPÍTULO 2: MARCO TEÓRICO.....	- 4 -
2.1. Big Data.....	- 4 -
2.1.1. <i>Definición de Big Data</i>	- 4 -
2.1.2. <i>Características de Big Data</i>	- 5 -
2.1.3. <i>Ventajas de Big Data</i>	- 10 -
2.1.4. <i>Desventajas de Big Data</i>	- 12 -
2.2. Empresas de comercialización masiva.....	- 14 -
2.2.1. <i>Definición de Comercialización Masiva</i>	- 14 -
2.2.2. <i>Características de las Empresas de Comercialización Masiva</i>	- 15 -
2.2.3. <i>Modo de Operación en el Mercado de las empresas de Comercialización Masiva</i>	- 17 -
3. CAPÍTULO 3: BUSINESS INTELLIGENCE Y GENERACIÓN DE CONOCIMIENTO.....	- 21 -
3.1. Definición de Business Intelligence.....	- 21 -
3.2. Características de Business Intelligence.....	- 22 -
3.3. Relación con la Generación de Conocimiento.....	- 26 -
3.4. Objetivos de Business Intelligence.....	- 30 -
3.5. Funciones Principales de Business Intelligence.....	- 33 -
4. CAPÍTULO 4: COMPARACIÓN DE LAS HERRAMIENTAS PARA LA GESTIÓN DE BIG DATA.....	- 37 -
4.1. Herramientas de Software Libre.....	- 37 -
4.2. Herramientas de Software Propietario.....	- 45 -
4.3. Análisis Comparativo.....	- 52 -
4.3.1. <i>Análisis Comparativo Funcional</i>	- 54 -
4.3.2. <i>Análisis Comparativo en Generación de Reportes</i>	- 60 -
4.3.3. <i>Análisis Comparativo en Usabilidad</i>	- 63 -
4.4. Recomendaciones técnicas para empresas de comercialización masiva ...	- 67 -

5. CAPÍTULO 5: CONCLUSIONES, RECOMENDACIONES Y BIBLIOGRAFÍA	-
68 -	
5.1. Conclusiones	- 68 -
5.2. Recomendaciones	- 69 -
5.3. Bibliografía	- 71 -
ANEXOS	- 75 -
Anexo 1: Glosario de siglas y términos técnicos	- 75 -

1. CAPÍTULO 1: FORMULACIÓN DEL PROBLEMA

En este capítulo se describirán los objetivos del presente trabajo de disertación, así como también el problema a analizar junto con sus antecedentes y justificación respectivos.

1.1. Antecedentes

Gracias a la evolución de la tecnología junto con el intercambio masivo de información, hoy en día se pueden evidenciar grandes volúmenes de datos que van en constante crecimiento, razón por la cual es necesario uno o varios sistemas de almacenamiento más sofisticados.

El ser humano justamente es uno de los personajes principales dentro de este contexto, puesto que cada día se encuentra manipulando, almacenando y gestionando todo tipo de información, acompañada del gran avance tecnológico y de nuevos desafíos que suponen analizar, descubrir y sobre todo entender un poco más allá de lo que las herramientas tradicionales pueden brindar.

Justamente, dentro de las herramientas disponibles para lo que es el manejo de Big Data se encuentran las de software libre y propietario, que son utilizadas por todo tipo de empresas para el manejo correspondiente de información. Sin embargo muchas de ellas no conocen los beneficios, ventajas y desventajas que estas herramientas ofrecen en cuanto a costos, funcionamiento y gestión de la información.

La comercialización masiva en los últimos tiempos ha venido tomando fuerza, ocupándose de lo que los clientes desean y también sirviendo de guía para lo que se produce y lo que se ofrece. Dentro de este proceso social también se destaca el gran manejo de información en lo que tiene que ver con productos y servicios para la venta al consumidor.

Es por todo esto que la razón principal para llevar a cabo esta investigación es analizar qué tipo de herramientas se ajustan mejor a las empresas de comercialización masiva, junto con la gestión de Big Data para el manejo y explotación adecuada de la información.

1.2. Justificación

En la actualidad, organizaciones y empresas han optado por la adquisición de herramientas de software libre y propietario orientadas a Big Data para poder solucionar problemas de manejo, gestión y almacenamiento de información.

Para justificar el desarrollo del presente proyecto de disertación de grado, se realizará un análisis comparativo entre estos dos grupos o categorías de herramientas orientadas a las empresas de comercialización masiva, con el afán de descubrir qué tipo de herramientas serían las más adecuadas para el entorno de trabajo en el que se encuentran y la gran cantidad de información que manejan. Además de brindar los mencionados beneficios, la presente investigación ayudará a resolver problemas dentro del contexto a analizar, complementando con nuevos conocimientos para finalmente obtener un mejor manejo de la información. En particular el problema observado nace cuando una empresa de comercialización masiva intenta mejorar su productividad a través de la integración de un sistema de Big Data propietario. Al dominar la herramienta en el mediano plazo se evidencian las deficiencias del software debido a las limitaciones propias de un sistema pagado, obligando a este tipo de empresas a migrar hacia plataformas de software libre o disminuir su crecimiento debido al uso de estas herramientas.

En lo personal, la manera de contribuir es reforzando conocimientos ya existentes conjuntamente con los nuevos que se obtendrán conforme se vaya desarrollando el presente trabajo.

1.3. Objetivos

1.3.1. *Objetivo General*

Realizar un análisis comparativo de herramientas de software libre y propietario para la gestión de Big Data, aplicado a empresas de comercialización masiva.

1.3.2. *Objetivos Específicos*

- Investigar y explicar el funcionamiento de las empresas de comercialización masiva.
- Investigar sobre conceptos relacionados con Big Data.
- Identificar beneficios, ventajas y desventajas que ofrecen las herramientas de software libre y propietario dentro del contexto de Big Data.
- Analizar aspectos relacionados con Business Intelligence y generación de conocimiento.
- Comparar, analizar y definir una solución ajustada a la realidad de las empresas de comercialización masiva.

1.4. Alcance

El presente proyecto de disertación de grado, culminará con la entrega de un documento que contenga el análisis de herramientas de software libre y propietario para la gestión de Big Data, así como la herramienta recomendada para empresas de comercialización masiva.

2. CAPÍTULO 2: MARCO TEÓRICO

En este capítulo se describirán los conceptos más importantes sobre Big Data y Comercialización Masiva, temas principales en los que se basa el fundamento teórico del presente trabajo.

2.1. Big Data

Existen varios criterios que describen lo que significa Big Data, entre los cuales se citan las definiciones más importantes y relevantes, así como también sus características, ventajas y desventajas.

2.1.1. Definición de Big Data

Viktor Mayer-Schönberger y Kenneth Cukier (2013) definen Big Data como “La capacidad de la sociedad para aprovechar la información en nuevas formas de producir ideas útiles o bienes y servicios de valor significativo” (p.2).

En un artículo publicado por Ricardo Barranco Fragoso (2012), miembro del grupo de software de IBM¹ en México, la definición de Big Data plantea lo siguiente: “Es la tendencia en el avance de la tecnología que ha abierto las puertas hacia un nuevo enfoque de entendimiento y toma de decisiones, la cual es utilizada para describir enormes cantidades de datos que tomaría demasiado tiempo y sería muy costoso cargarlos a un base de datos relacional para su análisis”.

La investigadora Gartner² también hace referencia a lo que es Big Data, e indica que “Big Data es gran volumen, alta velocidad y alta variedad de activos o bienes

¹ IBM – Empresa multinacional estadounidense de tecnología y consultoría con sede en Armonk, Nueva York.

² Gartner – empresa consultora y de investigación de las tecnologías de la información con sede en Stanford, Connecticut, Estados Unidos.

de información que exigen formas innovadoras y rentables de procesamiento de la misma para mejorar la comprensión y la toma de decisiones”.

Se puede mencionar en base a estas tres definiciones que Big Data es una metodología que permite gestionar y manejar gran cantidad de información, junto con distintas herramientas y sobre todo alta capacidad de análisis y entendimiento para mejorar la toma de decisiones sobre distintos contextos relacionados con la tecnología.

Está claro que dichos contextos incluyen varias temáticas tales como dispositivos móviles, audio, video, sensores digitales, sistemas de geolocalización, redes sociales, entre muchas otras. Sin embargo, cabe destacar que los grandes protagonistas de toda esta moción de Big Data somos los seres humanos, puesto que cada día estamos manipulando toda clase de información y utilizándola para fines pertinentes.

Es por todo esto que Big Data se considera una revolución en el manejo y la gestión de la información, junto con su explotación en nuevos flujos y que permite adquirir una cultura de toma de decisiones.

2.1.2. Características de Big Data

La revista INFORMABTL³ en su artículo **5 características de Big Data** (2014), menciona las siguientes:

- **Volumen de los datos:** La información se produce en grandes cantidades y en períodos cortos de tiempo. Dicha información tiene que ver con las

³ INFORMABTL – medio especializado en estrategias de mercadotecnia *Below The Line* (uso de medios de comunicación no masivos, dirigidos o segmentados).

temáticas mencionadas en el punto anterior (redes sociales, internet, dispositivos móviles, etc.)

- **Velocidad de generación de los datos:** se habla de aproximadamente 5 exabites cada 10 minutos. Además se requiere de una velocidad de respuesta rápida, para obtener la información correcta en el momento preciso.
- **Variedad de los datos:** Dentro de este contexto existen tres clases de datos:
 - **Estructurados:** Son aquellos que tienen un formato y longitud definidos. Generalmente son almacenados en tablas, bases de datos organizadas y divididas de forma lógica. Ejemplos de este tipo de datos pueden ser las hojas de cálculo, plantillas, etc.
 - **Semiestructurados:** Son aquellos que siguen una especie de estructura implícita, pero no tan regular y contundente para ser gestionada como un dato estructurado. Ejemplos de este tipo de datos pueden ser las páginas web (HTML y XML), señales de tráfico, etc.
 - **No Estructurados:** Son aquellos que se muestran tal y como fueron recolectados, es decir que carecen de formato alguno. Ejemplos de este tipo de datos pueden ser correos electrónicos, foros, documentos de texto, documentos multimedia, etc.
- **Valor de los datos:** Qué parte de los datos que se generan es valiosa para lo que se trata de analizar o resolver.
- **Veracidad de los datos:** No toda la información generada es confiable, por lo que se hace necesario analizarla para sacar conclusiones posteriormente.

De las características antes mencionadas, vale la pena hacer énfasis en algunos puntos principales que son fundamentales para el análisis.

Si es que se habla del volumen de datos, está claro que éste va aumentando constantemente en el día a día. Sin embargo, es fundamental entender de donde proviene toda esa información, así como también la manera en la que va creciendo.

“ANÁLISIS COMPARATIVO DE HERRAMIENTAS DE SOFTWARE LIBRE Y PROPIETARIO PARA LA GESTIÓN DE BIG DATA EN EMPRESAS DE COMERCIALIZACIÓN MASIVA”

Como se comentaba en el anterior punto, los seres humanos almacenan y crean información en cantidades astronómicas. Big Data no se refiere a una cantidad en específico, pero es común encontrar cifras relacionadas con exabytes y petabytes.

Pero esas unidades no son las únicas que existen. El sitio web *Tipos de Computadora y Más Allá*⁴ en su artículo *Medidas de almacenamiento de la Información-Almacenamiento de Datos parte II* (2012) muestra en un gráfico las principales unidades de medidas de almacenamiento con sus respectivos símbolos y equivalencias. Dicho gráfico es el siguiente:

Unidades de Medidas de Almacenamiento			
Medida	Simbología	Equivalencia	Equivalente en Bytes
byte	b	8 bits	1 byte
kilobyte	Kb	1024 bytes	1 024 bytes
megabyte	MB	1024 KB	1 048 576 bytes
gigabyte	GB	1024 MB	1 073 741 824 bytes
terabyte	TB	1024 GB	1 099 511 627 776 bytes
Petabyte	PB	1024 TB	1 125 899 906 842 624 bytes
Exabyte	EB	1024 PB	1 152 921 504 606 846 976 bytes
Zetabyte	ZB	1024 EB	1 180 591 620 717 411 303 424 bytes
Yottabyte	YB	1024 ZB	1 208 925 819 614 629 174 706 176 bytes
Brontobyte	BB	1024 YB	1 237 940 039 285 380 274 899 124 224 bytes
Geopbyte	GB	1024 BB	1 267 650 600 228 229 401 496 703 205 376 bytes

Figura 2-01: Unidades de Medidas de Almacenamiento utilizadas en Big Data.
Fuente: Artículo *Medidas de almacenamiento de la Información-Almacenamiento de Datos parte II* (2012), publicado en el sitio web *Tipos de Computadora y Más Allá*.

Las empresas o compañías también contribuyen con la acumulación masiva de datos, puesto que manejan información acerca de clientes, proveedores, transacciones, operaciones, etc. Aquí es muy útil introducir el concepto de la velocidad de generación de los datos y relacionarlo con el tiempo de respuesta, dado que si se puede gestionar la información en tiempo real de manera correcta, dichas empresas o compañías adquieren una ventaja competitiva sobre las demás.

⁴ Blog creado por Carlos Olvera, Guanajuato, México.

Y como es de esperarse, esta ventaja competitiva se ve reflejada en la toma de decisiones sobre el contexto que se está analizando conjuntamente con la agilidad y el enfoque del manejo de información que cada vez va aumentando.

Está claro que la información proviene de distintas fuentes como por ejemplo las redes sociales y las actividades que se realizan diariamente con los smartphones. Millones de mensajes, tweets y publicaciones de todo tipo de datos hacen que la información que se maneja sea cada vez mayor en el día a día.

Además, es fundamental tener en cuenta el tipo de dato que se está obteniendo de acuerdo a la clasificación de variedad de los datos (estructurados, semiestructurados y no estructurados) mencionada anteriormente. Sin embargo, es muy recomendado enfocarse en la información que se debe analizar y el problema que se está tratando de resolver, con el fin de obtener un mejor panorama para recolectar la información que realmente se necesita.

Resulta un poco obvio pensar que la variedad de los datos tiene mucha relación con los conceptos de volumen y velocidad de generación de los mismos. Es decir que en cuanto mayor sea el volumen, mayor será la variedad de datos que se tengan que analizar y por ende mayor la velocidad de generación y crecimiento de los mismos.

De acuerdo a un artículo publicado por el sitio web DATAVERSITY⁵ denominado ***Big Data: 5 tipos de datos que se necesita explorar hoy*** (2012), una clasificación general de tipos de datos se detalla en el siguiente gráfico con el fin de entender mejor su representación:

⁵ DATAVERSITY - Revista especializada en línea que provee recursos educativos de alta calidad para negocios e información que tiene que ver con el manejo de datos.

Big Data Types

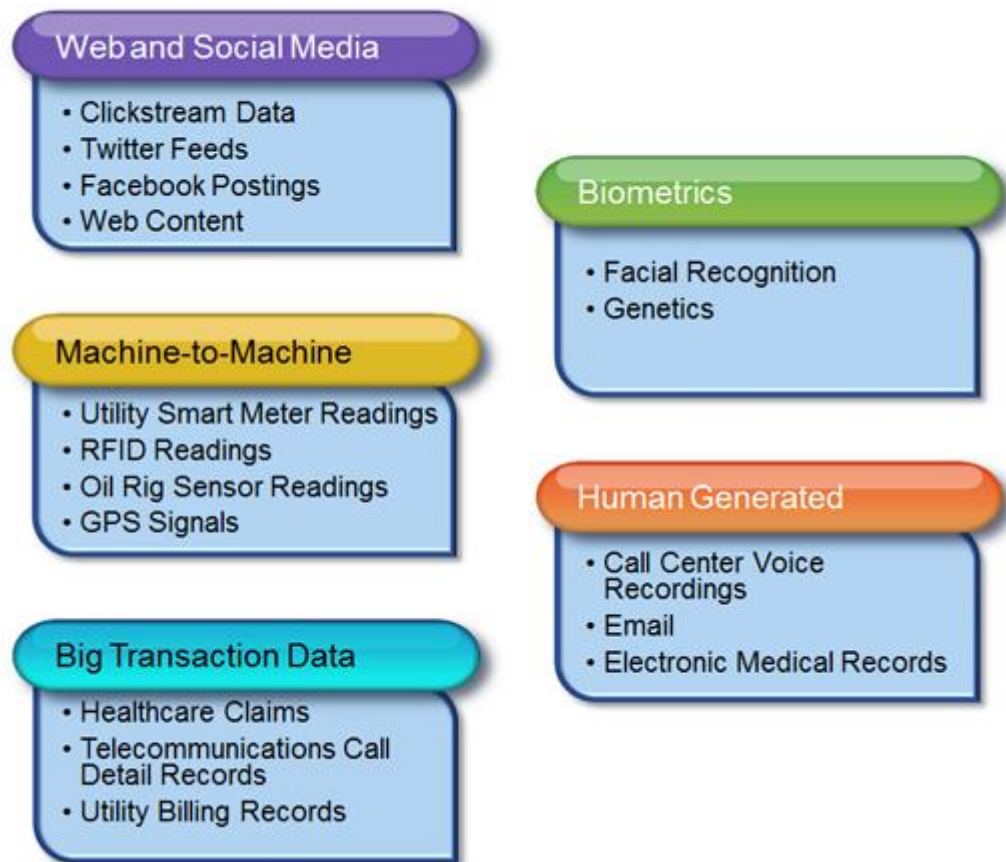


Figura 2-02: Tipos de Datos en Big Data. Fuente: Artículo *Big Data: 5 tipos de datos que se necesita explorar hoy* (2012), publicado por el sitio web DATAVERSITY.

A continuación se dará una breve explicación de cada uno de los tipos correspondientes a esta clasificación:

- 1. Web y Medios Sociales:** Tal y como se indica, aquí se maneja información correspondiente a redes sociales tales como Facebook, Twitter, LinkedIn, entre otras, así como también contenido web a manera de páginas, foros o blogs.
- 2. Máquina a máquina:** Medio de comunicación junto con otros dispositivos. Se utilizan sensores o medidores que captan algún evento en particular (velocidad, temperatura, presión, etc.) y lo transmiten a través de redes hacia otras aplicaciones que traducen dichos eventos en información útil y significativa.

3. **Grandes datos de transacción:** Tiene que ver con los registros de facturación de servicios públicos, reportes detallados de las llamadas en telecomunicaciones, quejas o reclamos, entre otros. Es importante resaltar que dichos datos transaccionales tienen formato tanto semiestructurado como no estructurado.
4. **Biometría:** Esta clasificación trata sobre seguridad, control y seguimiento en información biométrica, como por ejemplo reconocimiento facial, huellas digitales, genética, etc.
5. **Generados por los humanos:** Aquí se abarca toda la información que es generada por los seres humanos, tal como correos electrónicos, mensajes de texto, notas de voz, entrevistas, entre otros.

Cabe recalcar que en un futuro es muy probable que esta clasificación aumente gracias al avance de la tecnología y al crecimiento del volumen, velocidad de generación y variedad de los datos.

2.1.3. *Ventajas de Big Data*

El sitio web Siliconweek⁶ en su artículo *Las 5 ventajas competitivas que aporta Big Data* (2013) hace mención a las siguientes:

- Implementación de mejoras tecnológicas que posibilitan la adquisición de datos y permiten descubrir necesidades y puntos de mejora en la compañía
- El análisis de los datos puede mejorar sustancialmente la toma de decisiones dentro de una compañía reduciendo al mínimo los riesgos.
- Facilita que las compañías evalúen sus productos mediante el análisis de datos, obteniendo información muy valiosa que les permite crear nuevos productos o rediseñar los ya existentes.

⁶ Siliconweek – Portal de información tecnológica para los profesionales TIC

- Segmentación de los clientes para personalizar acciones. De esta manera las empresas pueden orientar sus servicios y satisfacer las necesidades de sus consumidores de forma específica.
- Mejora la accesibilidad y la fluidez de la información dentro de la propia empresa, creando una dinámica de trabajo más rápida y eficaz.

Sin embargo, también vale la pena recalcar que todas las ventajas antes mencionadas pueden generar a su vez una o varias ventajas competitivas a la compañía dentro de un entorno de gran distribución. Una ventaja competitiva es un cierto tipo de elementos o características que diferencian a una empresa de sus principales competidoras en el mercado, y esto es justamente lo que da el llamado valor agregado.

Otra potencial ventaja dentro de lo que es Big Data es que puede ayudar a la mejora y reducción de costos en productos, ya que se cuenta con mayor velocidad de desarrollo así como también mayor agilidad en las operaciones y control de datos que estos manejan. Por lo tanto los tiempos de desarrollo estimados se acortarán de una manera considerable.

Tampoco hay que olvidarse de que todo esto forma parte de un proceso y que las compañías no cosecharán todos estos beneficios a menos que sean capaces de gestionar todos los cambios de manera efectiva. No es solo el simple hecho de garantizar el éxito por tener más datos y en mejor calidad, sino hacer énfasis en determinadas áreas que contribuyen al mejor desempeño de la compañía utilizando Big Data.

Una de ellas es el liderazgo en equipos, con el fin de tener objetivos establecidos, un mejor entendimiento de cómo funciona el mercado, creatividad acompañada de

motivación y un buen trato con colegas y empleados. Es por esto que la presencia de uno o varios líderes es un punto clave para conseguir el éxito.

Los analistas de datos con habilidades para organizarlos en grandes cantidades también son una parte muy importante ya que de esta manera conocen su origen y el tipo para posteriormente adaptarlo al lenguaje o a la filosofía que maneja la compañía donde se están desempeñando. Son además personajes que pueden apoyar a los líderes reformulando alguna estrategia u objetivo según sea el caso.

Como no podía ser de otra manera, la tecnología está presente dentro de este contexto justamente con herramientas para analizar y gestionar la información, tanto de software libre como de software propietario. Es de gran utilidad combinar el uso de estas herramientas con posibles estrategias de Big Data y la integración de todas las fuentes de datos para poder alcanzar buenos y exitosos resultados.

Como se puede apreciar, todas las ventajas que nos ofrece Big Data en conjunto pueden llegar a formar una mejor cultura en todo el ámbito de gestión empresarial, detallado en los puntos que se han venido mencionando.

2.1.4. Desventajas de Big Data

Rado Kotorov en su artículo *Las desventajas de Big Data* (2013), publicado en el sitio web SG⁷ hace énfasis en los siguientes puntos:

- Aquellos que han decidido que su empresa dé el paso hacia objetivos mucho más ambiciosos, suelen minimizar el riesgo de lo que supone ese movimiento, es decir que no se toman en cuenta las necesidades reales ni

⁷ SG Software Guru – portal mexicano considerado un medio líder para los profesionistas de software en América Latina.

los recursos necesarios para afrontar un proyecto de Big Data, así como también los costos asociados.

- Uno se puede cuestionar indefinidamente si los beneficios están por encima de los riesgos. Pero hay que tener en claro que mientras más cuestiones se planteen, más arriesgado parecerá el proyecto.
- Suele conducir a aquellos que toman decisiones en una empresa a emplear una actitud pasiva en muchos aspectos. Por ejemplo, esperan que los competidores den el primer paso para después aprender de sus errores e implementas sus propios proyectos cuando llegue el momento adecuado.
- La mayoría de las compañías se empeñan en encontrar un caso de uso, y frecuentemente los más relevantes están relacionados con el análisis de datos que generan las máquinas. El problema con esto es que todavía son muy pocas las compañías que tienen la capacidad y equipamiento para generar esta información.

Frente a esta perspectiva se pueden mencionar otras desventajas, tales como la falta de profesionales capacitados y expertos en el tema de Big Data, la falta de inversiones destinadas hacia proyectos de esta índole, la dificultad para integrar tanto los procesos de la compañía como la información que se maneja y por último la calidad de los datos, en tanto y en cuanto éstos tengan valor, porque de lo contrario no merece la pena analizarlos.

Lamentablemente todo en la vida cotidiana tiene sus ventajas y desventajas. La cuestión es poder discernir, reflexionar y analizar si las ventajas que se tiene se encuentran por encima de las desventajas, como también los beneficios que se puede conseguir a futuro siendo consciente de los riesgos que esto conlleva.

Es por todo esto que la aplicación de Big Data en las compañías supone todo un reto que tiene como elementos principales la actitud y la responsabilidad de llevarla a cabo en una manera efectiva.

2.2. Empresas de comercialización masiva

En esta sección se mencionarán las definiciones y características más importantes, así como también el modo de operación en el mercado de las empresas de comercialización masiva, caso de estudio particular en el presente trabajo.

2.2.1. Definición de Comercialización Masiva

La comercialización masiva se define como “Un sistema que apunta vagamente a todo el mundo con la misma mezcla comercial. Este sistema supone que todo el mundo es igual. Comercialización masiva quiere decir tratar de vender a todo el mundo”. (Monografias.com, 2007)

El sitio web La Voz De Houston⁸ en su categoría pequeña y mediana empresa, hace mención en un artículo a la comercialización masiva de la siguiente manera: “Básicamente, en una estrategia de comercialización indiferenciada, los publicistas utilizan el mismo mensaje para todos los segmentos del mercado. Esto es similar a la comercialización masiva; los publicistas suelen crear un mensaje que atrae a todo el mundo, por lo que éste es a menudo general o simple para permitir que más personas puedan relacionarlo”.

Douglas Gray y Donald Cyr plantean la definición de comercialización masiva como “Un estilo de comercialización en el cual el vendedor produce y distribuye en forma masiva un producto y procura atraer a todos para que lo adquieran” (p.133).

⁸ La Voz De Houston – portal web de noticias a nivel mundial.

Tomando como base las definiciones anteriormente mencionadas, se puede indicar que la comercialización masiva es una metodología para producir y vender productos sin especificar el perfil del mercado objetivo, ya que puede ser consumido por varios segmentos de la población. Justamente esto tiene que ver con la filosofía de atraer a “todo el mundo” sin considerar mayores detalles a fondo.

Dicho concepto aplicado a las empresas de comercialización masiva supone que los objetivos de dichas compañías se orienten o se encaminen a tal filosofía, principalmente en los procesos que tienen que ver con la venta y la producción.

Todo esto forma parte de la comercialización que puede considerarse un proceso social en el cual se intercambian actividades de venta y producción respectivamente entre todo tipo de compañías. Es por esto que las empresas de comercialización masiva vienen a ser un componente muy importante dentro de este gran contexto.

2.2.2. Características de las Empresas de Comercialización Masiva

El sitio web [Gestion.org](http://gestion.org)⁹ pone a consideración en un artículo relacionado con el marketing masivo los siguientes puntos:

- Se producen productos en forma masiva, por lo que su promoción también es masiva, su coste de producción es relativamente bajo y no tiene gran diferenciación.
- Su gran ventaja es que tiene un mercado potencial muy grande que puede afectar casi a cualquier persona, lo que hace que las necesidades de promoción no tengas por qué ser segmentadas.
- Las empresas de comercialización masiva producen, distribuyen y promocionan productos o servicios dirigidos hacia todas las personas.

⁹ Gestion.org – portal web sobre gestión empresarial

Siguiendo con la misma perspectiva, es importante mencionar el aporte que realiza el Licenciado Diego Vázquez (2011) en su artículo sobre Mercados de Consumo, en el cual menciona algunas características que destacan a las empresas de comercialización masiva:

- El acceso a los productos comercializados en este segmento debe ser masivo, es decir, que una gran masa de personas puede y debe tener acceso a la compra de los mismos. Al ser masivo, el producto no distingue en ser demandado por alguna franja geográfica o social determinada.
- Cuando un producto está en esta clasificación por lo general se considera la “P” de plaza como la más importante, debido a que si el consumidor no encuentra el producto en determinado local este no tendrá drama de buscar en otro punto de venta.
- Finalmente, el criterio más importante para determinar si un mercado de consumo es masivo, es analizar si en este mercado se cumplen alguna de las necesidades descritas en la Pirámide Maslow (gráfico adjunto). En esta clasificación los consumidores no están dispuestos en hacer grandes esfuerzos en sus compras, debido a que no pasan por muchos procesos de análisis de los productos.



Figura 2-03: Pirámide de Maslow. Fuente: Artículo *Mercados de Consumo* publicado por el Licenciado Diego Vázquez (2011)

En resumen, se puede comentar que la característica más importante de las empresas de comercialización masiva es que se dirigen a un grupo de consumidores en masa, donde se tiene mucha oportunidad de venta, fabricación y distribución. No tienen ningún tipo de segmentación o categorización en el mercado y la comercialización se relaciona con algún tipo de necesidad por parte de los consumidores.

Cabe mencionar que también se caracterizan por el tipo de productos que comercializan (pueden ser específicos o diversificados), y a su vez en la eficacia o agilidad del servicio que brindan.

Aunque el tipo de productos que se ofrecen sea distinto, no hay que olvidarse de que la única filosofía de las empresas de comercialización masiva es la de dirigirse a todo el mundo.

2.2.3. Modo de Operación en el Mercado de las empresas de Comercialización Masiva

Es lógico pensar que una empresa debe tener ciertas áreas debidamente formadas y estructuradas para su correcto desempeño y funcionamiento. Partiendo de esta aclaración, cualquier compañía puede adaptar dichas áreas hacia su filosofía de mercado y las estrategias que se manejan junto con la misma.

Para Cecilia Caferri, experta en negocios, existen 7 áreas básicas para operar una empresa, las mismas que han sido expuestas en un artículo sobre Negocios dentro del sitio web About.com¹⁰. Dichas áreas son las siguientes:

¹⁰ About.com – La mayor fuente de contenidos de expertos en Internet que ayuda a los usuarios a responder preguntas, resolver problemas y aprender algo nuevo.

- **Área de Gestión:** Es fundamental para conducir el negocio, ya que tiene por función la dirección de la empresa por la cual una o varias personas tratan de lograr los objetivos que la organización se ha marcado.
- **Área de Recursos Humanos:** Es el área encargada de seleccionar, contratar, remunerar y capacitar al personal de la empresa. La actividad humana es un elemento vital del que no se puede prescindir en el mundo de la compañía.
- **Área de Comercialización (Marketing):** Dentro de esta área se incluyen las actividades de marketing para conocer las necesidades del mercado, determinar productos y servicios que satisfacen las necesidades de los consumidores potenciales, fijar precios, promover y distribuir.
- **Área de Contabilidad y Finanzas:** La contabilidad y finanzas son un conjunto de técnicas que se utilizan para registrar la información cuantitativa expresada en unidades monetarias de las transacciones que realiza una entidad económica, con el objetivo de facilitar la toma de decisiones en relación con dicha entidad económica.
- **Área de Producción:** Incluye todo el conjunto de actividades a través de las cuales la empresa crea los productos o presta los servicios que son el objeto de su actividad.
- **Área de Administración:** Es el área encargada de controlar toda la documentación de la empresa.
- **Área de Aspectos Legales:** Es el marco que regula las acciones de las empresas, tales como contratos y otras exigencias legales y establece los impuestos que éstas deben pagar al estado.

El modo de operación de una empresa común y corriente viene sustentado en cada una de las áreas descritas anteriormente. Es necesario ejecutar las acciones correspondientes en todas ellas para desempeñar un funcionamiento adecuado dentro del mercado.

Pero la pregunta es: ¿Qué tienen que ver éstas áreas con las empresas de comercialización masiva? Y eso es justamente lo que se quiere debatir y relacionar a continuación.

Dentro de las empresas de comercialización masiva, es necesario destacar algunos elementos importantes que forman parte de su operación en el mercado y que son los siguientes:

- **Canales de Venta:** Es la manera en la que la empresa pone a disposición de los consumidores los productos que ofrece la misma. Dentro de este contexto se destacan la venta por Internet y el establecimiento de varias tiendas o locales comerciales que puedan satisfacer la necesidad del consumidor de una manera inmediata.
- **Logística:** Tiene que ver con la presencia de centros de distribución ubicados en zonas geográficas estratégicas que abastecen a las tiendas o locales comerciales descritos en el punto anterior de manera regular. El número de centros de distribución se puede relacionar directamente con el área geográfica dentro de la que pueden operar o atender.
- **Productos:** Por lo general, los productos ofrecidos no tienen variaciones, por lo que su elaboración es escasa. Es por esto que la actividad se resume en procesos de compra, abastecimiento, distribución y venta. La característica principal de los productos de venta masiva es la uniformidad.
- **Precios de Costo y Venta:** En esta sección el precio de costo y venta de los productos a ofrecer se ve modificado cada vez que se negocian nuevas condiciones de compra con los proveedores, a medida que van aumentando los volúmenes de compra.

En el siguiente gráfico se muestra la relación de los elementos fundamentales en la operación de las empresas de comercialización masiva en el mercado junto con las áreas más importantes para operar una empresa cualquiera, que se mencionaron al principio de esta sección.

“ANÁLISIS COMPARATIVO DE HERRAMIENTAS DE SOFTWARE LIBRE Y PROPIETARIO PARA LA GESTIÓN DE BIG DATA EN EMPRESAS DE COMERCIALIZACIÓN MASIVA”

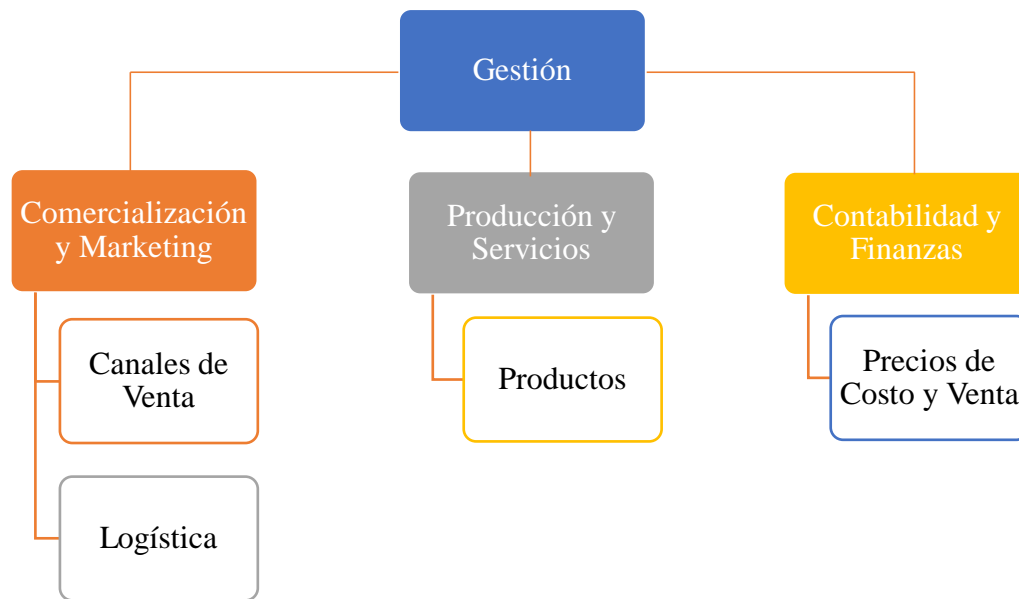


Figura 2-04: Relación de las áreas más importantes para operar una empresa junto con los elementos principales de la comercialización masiva

En este pequeño diagrama jerárquico no se han considerado las áreas de Administración, Recursos Humanos ni Aspectos Legales, dado que si bien son piezas clave en el funcionamiento de cualquier compañía, funcionan de manera más independiente y la relación con los elementos que se han mencionado como principales en las empresas de comercialización masiva no es de manera directa.

Resulta lógico suponer que el modo de operación en el mercado de las empresas de comercialización masiva dependerá exclusivamente del buen manejo de los elementos principales que se han discutido (Canales de Venta, Logística, Productos, Precios de Costo y Venta) junto con la colaboración de las áreas para operar una empresa cualquiera que tengan relación implícita con los mismos.

No hay que olvidar que todas las actividades que tengan que ver con el modo de operación de la compañía, deben estar alineadas a la planificación estratégica que normalmente es planteada por el área de Gestión, la misma que viene a ser el pilar fundamental de todo este proceso.

3. CAPÍTULO 3: BUSINESS INTELLIGENCE Y GENERACIÓN DE CONOCIMIENTO

En este capítulo se describirán los conceptos más importantes y relevantes acerca de Business Intelligence, así como también su relación con la generación de conocimiento, objetivos y funciones principales.

3.1. Definición de Business Intelligence

Sabherwal (2007, 2008) define Business Intelligence como “La provisión de toma de decisiones con información y conocimientos valiosos mediante el aprovechamiento de una variedad de fuentes de datos, así como información estructurada y no estructurada” (p.6).

Josep Curto Díaz (2012) también hace referencia al significado de Business Intelligence, indicando lo siguiente: “Se entiende por Business Intelligence al conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización” (p.18).

La investigadora Gartner tiene su propia definición de Business Intelligence, la cual menciona que “Business Intelligence es un término genérico que incluye aplicaciones, infraestructura, herramientas y las mejores prácticas que permitan el acceso y el análisis de la información para mejorar y optimizar las decisiones junto con el rendimiento”.

En base a las tres definiciones anteriores, se puede inferir que Business Intelligence es un proceso que involucra ciertas técnicas, metodologías y herramientas para transformar los datos en información significativa y útil para la organización, mejorando y optimizando la toma de decisiones dentro de la misma.

Business Intelligence se utiliza para comprender las capacidades con las que cuenta una determinada organización y sobre todo para responder de manera eficiente a las problemáticas que vayan apareciendo, como por ejemplo la tecnología, el entorno actual de competencia, tendencias y direcciones futuras de los mercados, entre otras.

Por lo explicado anteriormente, se puede deducir que Business Intelligence es una perspectiva que ofrece una ventaja competitiva dentro del mercado, tanto en operaciones simples como complejas que tengan el objetivo de satisfacer las necesidades de la empresa o compañía, basadas en la toma de decisiones, análisis y explotación de la información.

3.2. Características de Business Intelligence

Edison Medina La Plata (2012) hace énfasis en algunos elementos de suma importancia que caracterizan al contexto de Business Intelligence, y que son los siguientes:

- **Integración y universalización de la información:** Esto tiene que ser aprovechado para dar respuesta a las necesidades analíticas. Es decir que todos los departamentos de la empresa proporcionarán información y al mismo tiempo se beneficiarán de ella, teniendo acceso a cualquier tipo de información que sea de su utilidad y que puede provenir de otros departamentos.
- **Conversión de datos en información valiosa:** La empresa puede producir gran cantidad de datos que pueden estar a disposición de los diversos usuarios. Sin embargo es fundamental la obtención de información que sea de utilidad para algún departamento en base a comportamientos o indicadores de desempeño.
- **Apoyo en un conjunto de herramientas para el análisis de datos:** Dichas herramientas facilitan la extracción, la depuración, el análisis y el almacenamiento de los datos generados en una organización, con el fin de apoyar la toma de decisiones dentro de la misma.
- **Capacidades de análisis multidimensional:** Este análisis permite navegar y profundizar a través de los datos, agregando o desagregando de acuerdo a las

dimensiones que se consideren más importantes, permitiendo también el seguimiento de los indicadores de negocio más relevantes de la empresa.

Ya se ha mencionado que uno de los puntos fuertes de Business Intelligence es la gestión de la información con el fin de optimizarla para mejorar la toma de decisiones en base a resultados concretos. Sin embargo hay que recalcar que dentro de una organización existen diferentes tipos de necesidades de información que se manejan de acuerdo a un contexto operacional determinado.

El siguiente diagrama ilustra de mejor manera lo que se refiere a las necesidades de la información en la empresa:

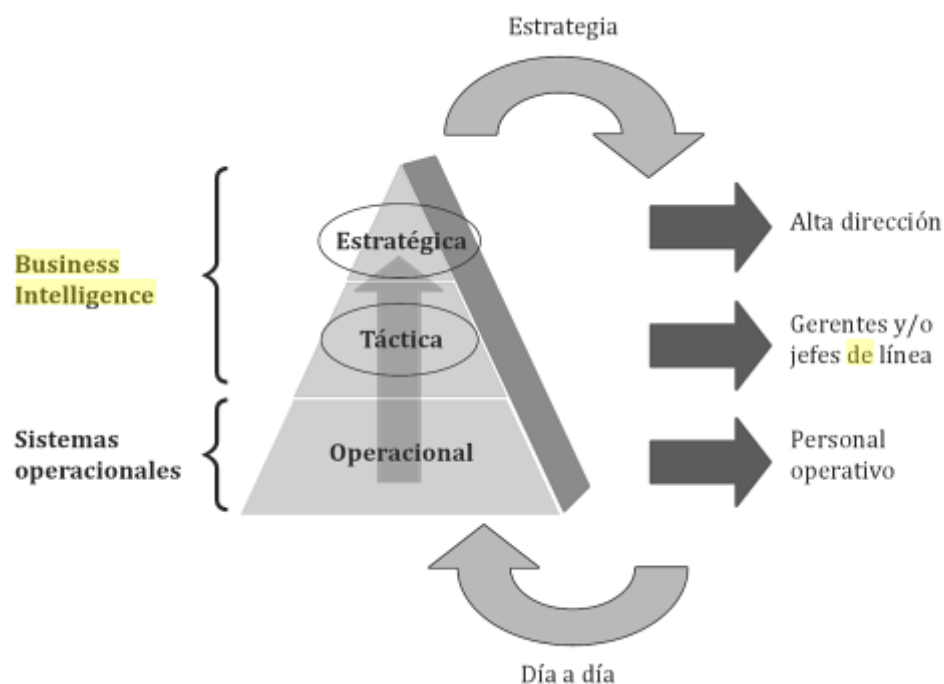


Figura 3-01: Necesidades de Información en la Empresa. Fuente: *Business Intelligence, una guía práctica*. Segunda edición. UPC. Por Edison Medina La Plata (2012) (p.29)

“ANÁLISIS COMPARATIVO DE HERRAMIENTAS DE SOFTWARE LIBRE Y PROPIETARIO PARA LA GESTIÓN DE BIG DATA EN EMPRESAS DE COMERCIALIZACIÓN MASIVA”

Dentro de dicho entorno, los sistemas de Business Intelligence toman los datos que provienen que cualquier transacción realizada en el nivel operacional para convertirlos en información ordenada, integrada y homogénea para la gestión correspondiente en los niveles táctico y estratégico. Por lo tanto toda la información que se vaya generando dentro de la organización es compartida hacia todos los niveles de gestión en la misma, donde los usuarios de cada nivel tienen la responsabilidad de manejar dicha información de acuerdo a las actividades que desempeñan.

También merece la pena hablar un poco sobre los componentes que caracterizan a un sistema de Business Intelligence. El siguiente gráfico ilustra la arquitectura más sencilla de una de estas soluciones:

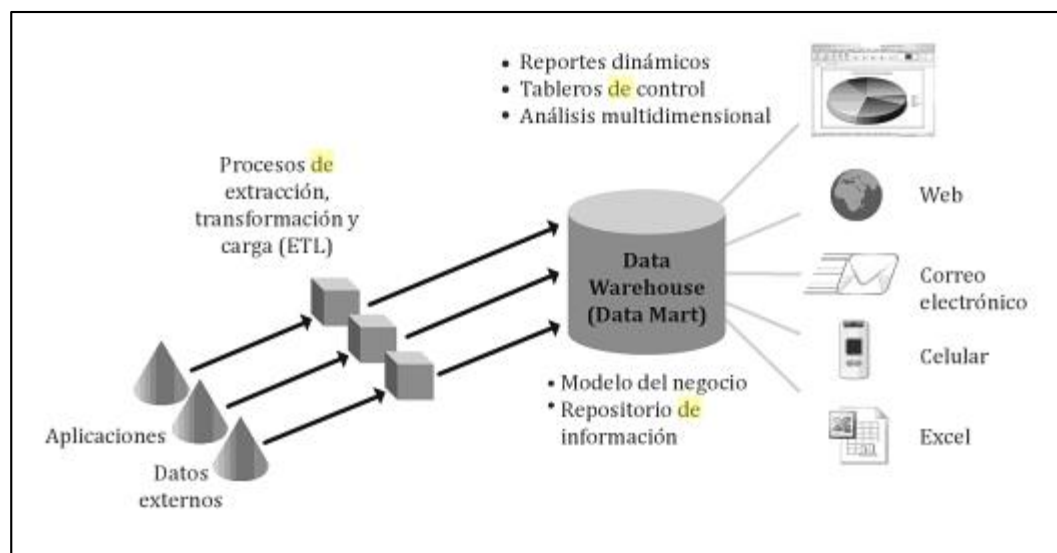


Figura 3-02: Arquitectura de una solución de Business Intelligence. Fuente: *Business Intelligence, una guía práctica*. Segunda edición. UPC. Por Edison Medina La Plata (2012) (p.31)

La interpretación del gráfico anterior se sigue de la siguiente manera:

La parte de aplicaciones y datos externos tiene que ver específicamente con la recolección de información de la que ya se habló unos pocos renglones más arriba. Dichos datos están preparados para ser almacenados y analizados posteriormente por las herramientas de extracción, transformación y carga (ETL).

Dichas herramientas lo que hacen es extraer los datos y convertirlos en un nuevo formato para el proceso de transformación. Como es lógico pensar, previamente se realiza un análisis para verificar si los datos cumplen con el formato esperado, casi contrario se rechazan o se eliminan. En el proceso de transformación se aplican ciertas reglas o normas de acuerdo a la política de negocio sobre los datos extraídos para así convertirlos en datos listos para ser cargados al Data Warehouse.

El Data Warehouse es un repositorio de datos a nivel de toda la organización. El objetivo de este es proporcionar información por diversos métodos a cualquier nivel que la requiera de manera general. Sin embargo, si se quiere contar con la información de cierta área o áreas específicas del negocio, existen los Data Mart que hacen posible dicha funcionalidad.

Al final, la idea es obtener información valiosa que pueda ser enviada al personal que la requiera mediante correo electrónico o dispositivos móviles. Con todo este proceso se está haciendo referencia a la primera característica en lo que tiene que ver con integración y universalización de la información, donde lo que se busca es la producción de información y el beneficio correspondiente con los resultados obtenidos.

Se puede inferir que el Data Warehouse además es un repositorio que sirve de apoyo para el negocio pero sobre todo es un soporte para el proceso de toma de decisiones. ¿Por qué? Pues porque permite el análisis de la información desde diferentes perspectivas o puntos de vista y con gran velocidad de respuesta.

Y es por esta razón que vale la pena destacar al Data Warehouse como una característica o un componente sumamente importante en los sistemas de Business Intelligence.

Resumiendo, existen 3 puntos fundamentales en los que se basan las características de los sistemas de Business Intelligence y que son los siguientes:



Figura 3-03: Elementos principales que caracterizan a los sistemas de Business Intelligence

En conclusión, todas estas características de los sistemas de Business Intelligence apoyan a la organización en la toma de decisiones y en la mejora de su rendimiento a partir de parámetros reales que sirven como punto de partida para la generación de conocimiento, situación que se explicará con mayor detalle en el siguiente punto.

3.3. Relación con la Generación de Conocimiento

Se ha venido hablando mucho de lo que son datos e información, ya que son elementos de vital importancia que forman parte de todo el contexto de Business Intelligence. Sin embargo existe otro elemento que es necesario destacar y del que dependen tanto la toma de decisiones como el rendimiento en la organización.

¿Cuál viene a ser ese elemento? Pues es justamente el CONOCIMIENTO, ya que a partir de él se deriva el respaldo, las acciones y el camino a seguir frente a alguno o varios acontecimientos que sean de importancia considerable para la organización. Dicho esto, el conocimiento puede considerarse tranquilamente como un parámetro que sirve de guía para la anticipación a tales acontecimientos luego de un análisis plenamente ejecutado sobre los datos correspondientes que provienen del ambiente operacional en el día a día.

Pero también es necesario entender de donde proviene tal conocimiento y desde donde se obtienen las bases o parámetros necesarios para su posterior generación. Pues bien, el conocimiento tal y como se lo trata en esta sección es un componente principal que forma parte de los sistemas de Business Intelligence y que tiene su lugar específicamente en la denominada “Pirámide del Conocimiento o Pirámide DIKW”.

Para explicar de mejor manera dicho modelo piramidal, merece la pena mencionar el siguiente apartado junto con el gráfico correspondiente:

José Juan Mora Pérez (2012) hace la siguiente referencia con respecto al modelo DIKW: “DIKW es el acrónimo en Inglés de Datos, Información, Conocimiento y Sabiduría y se trata de un modelo en forma de pirámide que permite estructurar la relación entre los elementos citados previamente”. El gráfico siguiente muestra exactamente dicho modelo con los elementos clasificados respectivamente:

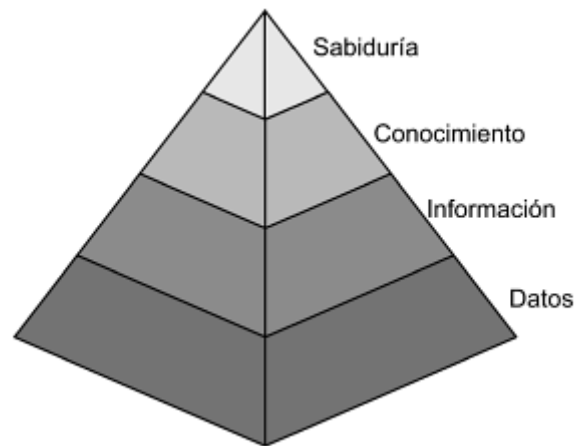


Figura 3-04: Jerarquía DIKW. Fuente: *Capacity Planning IT, una aproximación práctica*. Por José Juan Mora Pérez (2012) (p.473)

Tal y como el gráfico lo indica, el conocimiento se encuentra en un tercer nivel, inmediatamente después de la información y antes de la sabiduría. De aquí se puede inferir que mientras mayor sea el nivel en la pirámide, se adquiere mayor conocimiento sobre algún tema en específico, lo que posteriormente le da un mayor significado al tema o temas sobre los que se quiera tratar.

El conocimiento también está estrechamente ligado en el proceso de toma de decisiones para tratar de dar cumplimiento a los objetivos y finalidades que la organización maneja. Ahora la pregunta es: ¿De qué manera se puede llegar a tal nivel de conocimiento para tomar las decisiones que mejor se ajusten al entorno empresarial de la organización?

Para responder a dicha pregunta, hay que regresar por un momento al punto anterior de este capítulo, donde se citó específicamente el diagrama de necesidades de información que tiene una empresa. ¿Por qué justo allí? A continuación la explicación:

Está claro que para llegar al conocimiento primero hay que pasar por los datos y después convertirlos en información, para posteriormente añadirle el valor y la experiencia necesarios con el fin de alcanzar tal conocimiento. Se había mencionado también que

los datos se comienzan a recolectar en el nivel operacional y que debido al orden y el modo que estos sean interpretados y analizados, van adquiriendo mayor valor para poder ser gestionados en los niveles táctico y estratégico respectivamente, siendo estos los niveles superiores en aquella pirámide.

Y es por eso que resulta lógico realizar una comparación entre la pirámide de las necesidades de la información en una empresa y la jerarquía DIKW, donde la relación se daría de la siguiente manera:

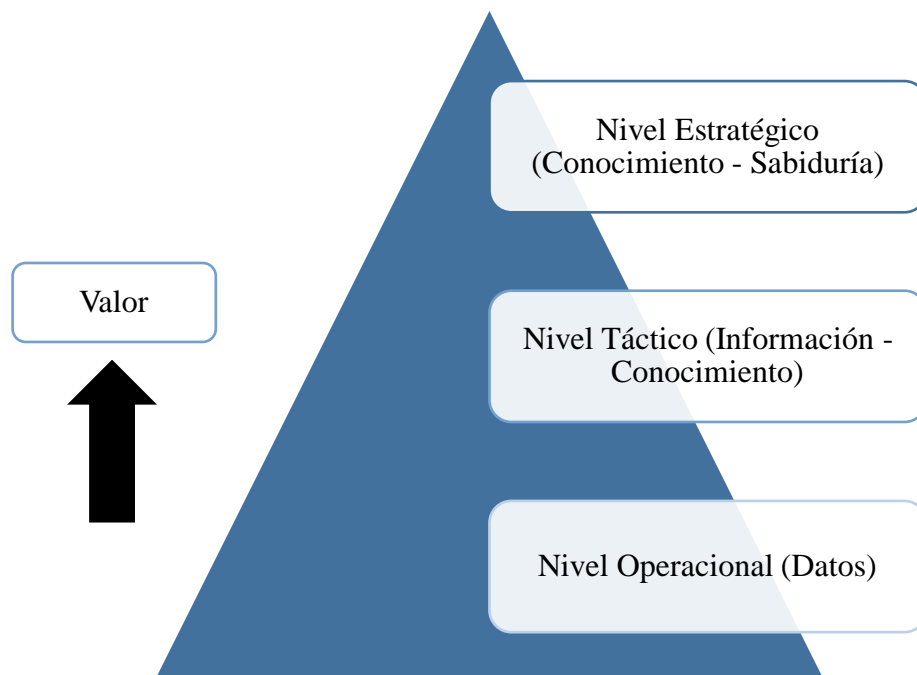


Figura 3-05: Relación entre las necesidades de información y la jerarquía DIKW

La interpretación de este gráfico es sencilla. Se había mencionado que todo parte desde el nivel operacional con la recolección de datos y que cada vez que se aplique un análisis adecuado, ésta será la puerta de entrada hacia el siguiente nivel relacionando las dos categorías (Necesidades de información y jerarquía DIKW).

Es obvio que para avanzar de nivel, el significado y el valor de los datos que se manejan al principio va en aumento, y también resulta lógico afirmar que los niveles de dichas categorías van a la par. Es por esta razón que el conocimiento se encuentra en el nivel táctico y estratégico, dado que es justo allí donde interviene el proceso de toma de decisiones principalmente en las áreas administrativas y en la alta dirección con el fin de cumplir con los objetivos estratégicos a los que está orientada la organización.

En conclusión, la relación que tienen los sistemas de Business Intelligence con la generación de conocimiento es directa y fundamental para la toma de decisiones dentro de la organización, identificación de posibles tendencias en el comportamiento de los clientes, anticipación a los acontecimientos que pueden producir un cambio en el entorno empresarial, monitorización de resultados y estrategias, entre otras situaciones.

3.4. Objetivos de Business Intelligence

Hasta el momento, se han mencionado algunas características de los sistemas de Business Intelligence que ponían como elementos principales al manejo y gestión de la información, así como también a la toma de decisiones. De igual manera se ha manifestado la importancia del conocimiento como factor fundamental dentro de todo este proceso.

Al rato de describir los objetivos, se puede pensar que los mismos indican que es lo que se pretende o que es lo que se busca establecer con esta idea de Business Intelligence. Bajo esa premisa, es necesario recurrir al siguiente apartado para posteriormente deducir las perspectivas que más se ajusten a la descripción que se quiere llegar en cuanto a los objetivos se refiere.

Dicho apartado hace referencia a lo siguiente:

Josep Curto Díaz (2012) indica varios aspectos que destacan en los sistemas de Business Intelligence y que bien podrían tomarse como objetivos potenciales de los mismos. Tales aspectos son los siguientes:

- **Crear un círculo virtuoso de la información:** Esto tiene que ver justamente con el proceso de transformación de los datos en información, conocimiento y sabiduría, que permite tomar mejores decisiones que se traducen en mejores resultados y por ende en la generación de nuevos datos.
- **Permitir una visión más a detalle de la información:** Se refiere a una visión más única, histórica, persistente y de calidad con el fin de darle el uso más adecuado.
- **Manejar y mantener métricas dentro de la organización:** Entre estas métricas se pueden citar un par de ejemplos, tales como el caso de los indicadores claves de rendimiento e indicadores claves de metas que son fundamentales para la formulación de estrategias dentro de la empresa, como también para identificar los procesos claves de negocio acordes con la política del mismo.
- **Aportar información actualizada en todos los niveles:** No importa el nivel en el que se esté llevando a cabo algún tipo de gestión. Si no se entrega una información actualizada o en tiempo real, de nada servirá procesar esa información y peor aún llevarla a los niveles superiores.
- **Mejorar la comprensión de los sistemas de información dentro de la organización:** Lo que se pretende es proporcionar a los usuarios la información necesaria para que pueda ser analizada y gestionada en los distintos niveles mediante la tecnología. Además la idea es dar el apoyo suficiente a los usuarios para comprender los procesos operativos y gestiones que realizan en cada uno de los niveles.
- **Mejorar la competitividad de la organización en el mercado:** Este objetivo es uno de los predominantes, ya que se pretende que la empresa tenga mayor agilidad en lo que se refiere a la toma de decisiones, que pueda anticiparse a cualquier tipo de cambio y que tenga un acceso más rápido a la información.

Sin embargo, los elementos antes mencionados pueden resumirse de un modo más general o más global en el siguiente gráfico:

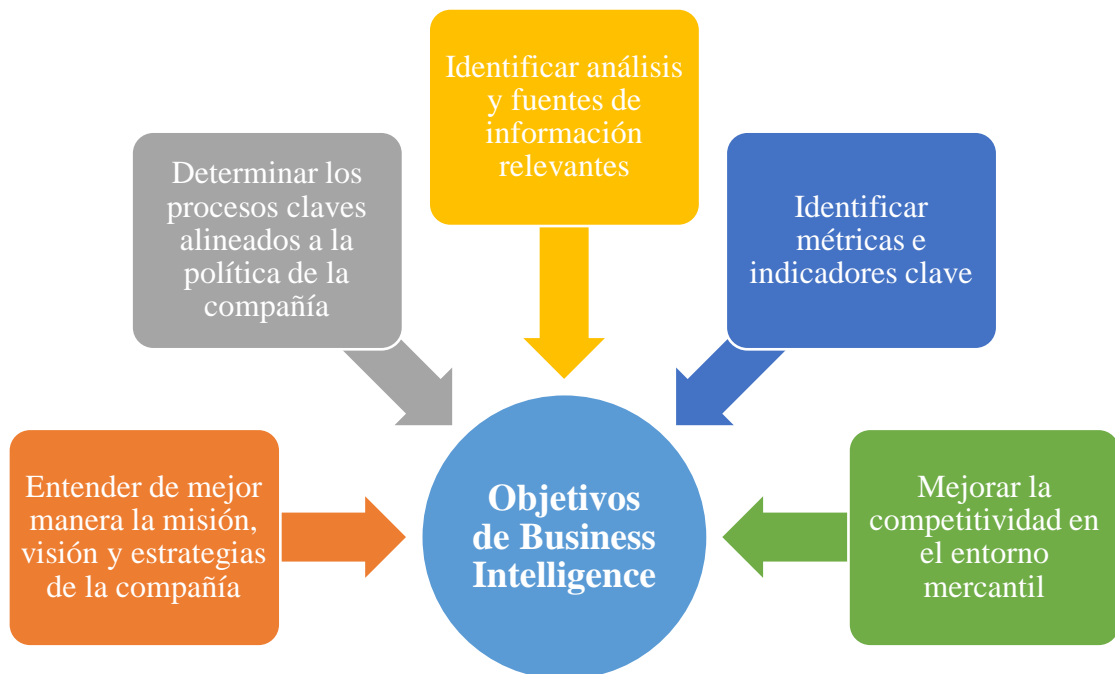


Figura 3-06: Objetivos de Business Intelligence en términos generales

De este modo se pueden apreciar los objetivos de Business Intelligence propuestos por Josep Curto Díaz en términos generales con el fin de plasmar un escenario ideal en donde las herramientas de Business Intelligence sean capaces de responder hacia tales objetivos.

Hay que tener en cuenta que los objetivos de Business Intelligence pretenden dar apoyo y soporte a las herramientas correspondientes dentro de la formulación de objetivos estratégicos, cuyo fin es brindar pautas o normas de actuación que van encaminadas a mejorar el rendimiento de la organización, generalmente a largo plazo.

3.5. Funciones Principales de Business Intelligence

Antes de iniciar con este punto, es necesario hacer la siguiente aclaración. Si bien el término de funciones principales puede resultar en cierto modo similar al de los objetivos, la diferencia es que como se comentó en el punto anterior los objetivos describen lo que se pretende o lo que se busca con los sistemas de Business Intelligence dentro de un escenario ideal. En cambio, las funciones principales describen exclusivamente lo que dichos sistemas hacen o realizan, independientemente del escenario en el que se encuentren.

Una vez realizada la aclaración para evitar alguna posible confusión, en el siguiente gráfico se hará referencia a algunas de las funciones principales de Business Intelligence propuestas por el Licenciado Ricardo Sánchez Montoya en su artículo sobre Inteligencia de Negocios (BI):



Figura 3-07: Funciones principales de Business Intelligence. Fuente: Artículo *Inteligencia de Negocios (BI)*, publicado por el Licenciado Ricardo Sánchez Montoya en el sitio web Monografías.com

Como se puede apreciar, las funciones principales de Business Intelligence operan independientemente del contexto en el que se manejen las herramientas que las aplican. A continuación se dará una breve explicación sobre cada una de ellas:

La generación de reportes globales o por secciones se encarga de analizar y reproducir la información requerida y transmitirla a los niveles que la necesiten. La idea es plasmar resultados sobre alguna temática y posteriormente analizarlos para ver su impacto a corto, mediano o largo plazo dentro de la organización.

Así mismo existen diferentes escenarios con respecto a una decisión. Esto depende del nivel de conocimiento que se tenga luego de que los datos hayan pasado por el debido proceso de análisis para convertirse en conocimiento. Aquí es justamente donde se tienen los escenarios a los que se hace mención, es decir las pautas o los caminos por donde se puede ir en base al proceso de toma de decisiones para el futuro de la organización.

Lo anterior también está ligado con proyecciones y pronósticos en diferentes áreas de interés, como se citó en el gráfico los ejemplos de ventas y devoluciones. Pero también se puede pronosticar comportamientos de los clientes en los últimos meses, compras y relación con los proveedores principales, proyecciones a futuro, entre otras situaciones que ya se han venido hablando a lo largo del capítulo.

Compartir información entre los departamentos es un elemento que ya se ha mencionado antes en la descripción de los niveles de la información que tiene una empresa, donde cada uno de los empleados se vuelve responsable de la información que maneja y también del uso que le dé a la misma. Esto permite también una visión más a detalle de la información, donde se pueden tener distintos criterios pero que persiguen un mismo fin.

La generación y procesamiento de los datos tampoco son términos nuevos en este punto, puesto que son componentes principales de los sistemas de Business Intelligence ya mencionados anteriormente. El análisis multidimensional permite realizar consultas de información que está distribuida en distintos ejes dentro de estructuras multidimensionales, así se puede navegar a través de los datos y obtener la información requerida según los parámetros que se hayan utilizado previamente al momento de realizar la consulta.

Finalmente, y como resulta obvio pensar, todo este contexto de los sistemas de Business Intelligence supone una mejora de servicio al cliente, dado que se les brinda una gran variedad de opciones o alternativas en los productos que ellos pueden necesitar.

También es necesario tomar en cuenta el proceso de funcionamiento básico de un sistema de Business Intelligence, y para esto considero necesario apoyarse en una Figura propuesta por Lee Wittschen (2004) y analizada por Elizabeth Del Carmen Zárate Gallardo (2013):

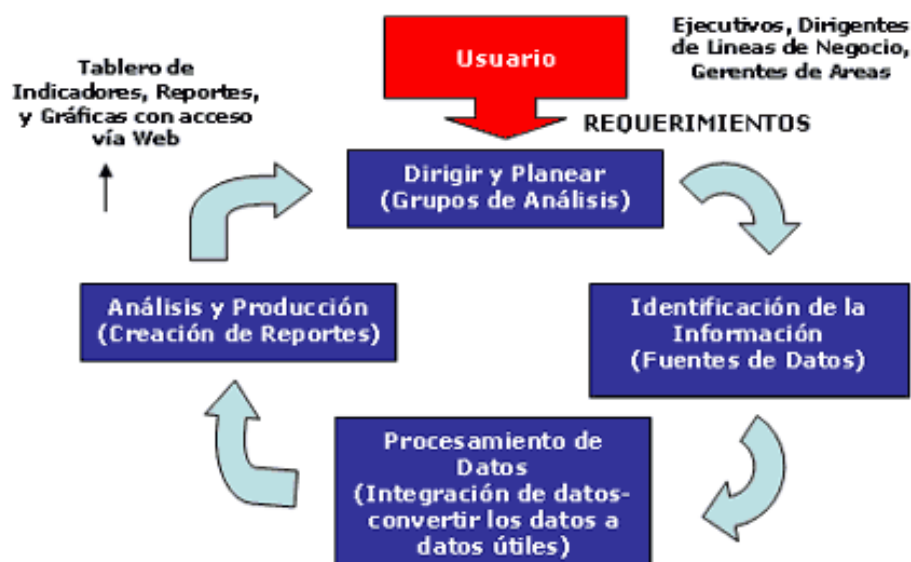


Figura 3-08: Proceso de Business Intelligence. Fuente: *Why Business Intelligence?* Por Lee Wittschen (2004), publicado en el artículo de Elizabeth Del Carmen Zárate Gallardo (2013) en el sitio web Gestipolis.com

A continuación el análisis del gráfico anterior propuesto por Zárate Gallardo, en su artículo *Inteligencia de Negocios* publicado en el sitio web Gestipolis.com¹¹:

- **Dirigir y Planear:** Esta en la etapa inicial en donde se definen los requerimientos e hipótesis para la formulación de objetivos con el fin de que cubran las necesidades principales en la organización.
- **Recolección de Información:** Aquí se identifica la información en base de datos para posteriormente tratar de obtener los resultados esperados.
- **Procesamiento de Datos:** Es importante integrar los datos dentro de un mismo formato para que puedan ser analizados, procesados e interpretados.
- **Análisis y Producción:** En esta fase se crean los reportes personalizados que proporcionan directrices o pautas de acción que serán importantes para la toma de decisiones.
- **Difusión:** Generación de los reportes personalizados con posibles interpretaciones y lineamientos para la gestión de las directrices o pautas de acción junto con la toma de decisiones.

Toda la parte conceptual que se ha tratado sobre el entorno de Business Intelligence y la generación de conocimiento sirve para entender de mejor manera su importancia dentro de la organización como una metodología generadora de beneficios y de ventaja competitiva para dicha organización, apoyándose en los objetivos y funciones principales previamente citadas y descritas.

¹¹ Gestipolis.com – portal web acerca de temas vinculados con la administración, la empresa y la economía.

4. CAPÍTULO 4: COMPARACIÓN DE LAS HERRAMIENTAS PARA LA GESTIÓN DE BIG DATA

En este capítulo se describirán las herramientas de software libre y propietario (una por cada caso) para posteriormente realizar el análisis comparativo en tres puntos importantes, los cuales son a nivel funcional, a nivel de generación de reportes y a nivel de usabilidad. Finalmente se mencionarán las recomendaciones técnicas para las empresas de comercialización masiva, caso de estudio en el presente trabajo de disertación.

4.1. Herramientas de Software Libre

En este punto se describirán las generalidades y los puntos más relevantes que poseen las herramientas de software libre. Sin embargo, antes de mencionar o citar la parte conceptual, es necesario hacerse la siguiente pregunta: ¿Qué se entiende por software libre?

Pues bien, se puede pensar que cuando se habla de software libre se habla de un software gratuito, sin restricciones y con libertad para ser usado en cualquier circunstancia donde dicho software pueda apoyar o solventar alguna necesidad en particular. Pero esta es una definición en términos generales, por lo tanto vale la pena explicar con un nivel más detallado cada uno de esos términos con el fin de exponer el concepto de software libre de una manera más clara y comprensible.

El movimiento de software libre encabezado por Richard Stallman¹², fundador de la *Free Software Foundation* (FSF) tiene como filosofía mantener público el código fuente de un programa de computador, de tal manera que cualquier persona pueda conocerlo, modificarlo, adaptarlo a sus necesidades específicas, copiarlo, etc.

¹² Proyecto GNU patrocinado por la Free Software Foundation (FSF) – URL: <http://www.gnu.org>

Según manifiesta Martín Carranza Torres (2004), “el principio básico del Proyecto GNU y de la *Free Software Foundation* es que el código fuente es fundamental para fomentar la informática y que la libre disponibilidad del código fuente es verdaderamente necesaria para que la innovación continúe”

Mientras tanto, el Dr. Santiago Acurio Del Pino¹³ (2014) en su artículo ***Protección al Software*** expone algunas características muy importantes acerca del software libre que son las siguientes:

“El software libre opera a través de una licencia pública general (GPL) que puede resumirse en 5 puntos: (p.18).

1. El código fuente debe ser revelado.
2. El uso y distribución del código fuente son libres.
3. El código fuente puede ser libremente modificado y copiado
4. Las modificaciones y mejoras que se introduzcan en el código fuente son también de uso libre
5. Los derechos sobre el código fuente no pueden ser restringidos por persona alguna”.

“El hecho de hablar de software libre, no significa que se esté hablando de software gratuito conocido como *freeware*, existen costos que pagar para cubrir gastos derivados de la protección a los derechos de autor. Esta tendencia se resume en dos puntos principales señalados por el mismo Stallman: (p.18).

- El software libre es una cuestión de libertad y no de precio.
- La cooperación es más importante que la Propiedad Intelectual”.

¹³ Abogado, Doctor en Jurisprudencia y Docente de la Pontificia Universidad Católica del Ecuador.

Dentro del mismo artículo (p.21) y también de acuerdo con la *Free Software Foundation*, software libre significa que los usuarios de un programa tienen las cuatro libertades fundamentales:

- Libertad 0: Libertad de usar el programa con cualquier propósito.
- Libertad 1: Libertad de estudiar el programa y adaptarlo a las necesidades propias.
- Libertad 2: Libertad de distribuir copias.
- Libertad 3: Libertad de mejorar el programa y hacer públicas las mejoras a los demás de modo que todos se beneficien

Dentro de esta perspectiva, se han mencionado varios puntos que a mi consideración son los más importantes para entender de mejor manera el significado de software libre con un enfoque hacia los usuarios y código fuente de los programas de computador.

Una vez que se ha hablado sobre el software libre, es momento de hablar sobre la herramienta que va a ser uno de los objetos de estudio en este capítulo. La herramienta elegida será *Pentaho Business Analytics* junto con el módulo *Pentaho Data Integration (PDI)*, dado que es una de las soluciones más populares en el mercado que integra tanto Big Data como Business Intelligence, temas fundamentales en los que se basa el presente trabajo.

El sitio web oficial de la compañía Pentaho¹⁴ en su sección *About Us* indica que “es una compañía líder en integración de datos y análisis de negocio con una plataforma de clase empresarial basada en código abierto para implementaciones diversas en lo que se refiere a Big Data”.

Vale la pena mencionar en el mismo artículo algunas de las razones más importantes por las que los líderes e innovadores eligen Pentaho:

¹⁴ URL: <http://www.pentaho.com>

- **Entrega de Datos gobernada:** Es definida como la capacidad de combinar datos confiables y oportunos de acuerdo al poder de análisis para todos los usuarios en todos los ambientes.
- **Fácil de encajar:** En cualquier entorno o aplicación de flujo de trabajo es sencillo encajar este tipo de soluciones ya sea en la nube o en dispositivos móviles.
- **Poder para acceder, preparar y mezclar los datos:** La plataforma entrega análisis precisos y listos a los usuarios finales de cualquier fuente de datos. Con el arrastre visual de los componentes se elimina la codificación y la complejidad, poniendo al alcance de los usuarios grandes volúmenes y fuentes de datos tradicionales en tiempo real.
- **Integración nativa con grandes fuentes de datos:** Pentaho permite convertir los grandes volúmenes de datos en grandes ideas, independientemente del tipo de análisis o del entorno de despliegue. La capa adaptativa de Big Data permite crear un acceso a prueba de futuro para todas las fuentes de datos con las que se esté trabajando.

Las razones citadas anteriormente muestran un escenario global de funcionamiento de las soluciones que brinda Pentaho. Sin embargo, las razones personales que me han llevado a elegir esta herramienta en términos generales son dos: popularidad y orientación a Big Data, que es una de las perspectivas principales ya mencionada en el capítulo 2.

Es importante aclarar que la compañía Pentaho ofrece dos soluciones que destacan dentro de este contexto y disponibles para descarga: *Pentaho Business Analytics* y *Pentaho Data Integration*, con la particularidad de que esta última viene integrada en la primera.

A continuación considero conveniente realizar un pequeño cuadro comparativo donde consta una pequeña descripción y algunas funciones principales de cada una de las

“ANÁLISIS COMPARATIVO DE HERRAMIENTAS DE SOFTWARE LIBRE Y PROPIETARIO PARA LA GESTIÓN DE BIG DATA EN EMPRESAS DE COMERCIALIZACIÓN MASIVA”

soluciones, según manifiestan Iñigo Castánder, Jon Kepa Gerrikagoitia y Aurkene Alzua-Sorzabal junto con la compañía Pentaho (2013):

Tabla 4-01: Comparativa entre las soluciones *Pentaho Data Integration* y *Pentaho Business Analytics*

SOLUCIÓN	DESCRIPCIÓN	FUNCIONES PRINCIPALES
Pentaho Data Integration (PDI)	Herramienta que facilita la obtención de datos de diversas fuentes. También facilita un ETL (extracción, transformación y carga) de alto rendimiento que soporta procesos de Big Data.	<ul style="list-style-type: none"> • Acceso a diferentes fuentes de datos. • Preparación para Big Data. • Generación del Data Warehouse. • Integración para la generación de informes, visualización y análisis de datos.
Pentaho Business Analytics	Entorno que permite analizar y generar diferentes tipos de visualizaciones, así como también la generación de informes y análisis predictivo.	<ul style="list-style-type: none"> • Integración de PDI • Mencionadas en la descripción.

Fuente: Estudio comparativo de las herramientas de Business Intelligence: Empoderando el criterio de selección a las PYMEs. Por Iñigo Castánder, Jon Kepa Gerrikagoitia y Aurkene Alzua-Sorzabal junto con la compañía Pentaho (2013).

Como se mencionó en algunas líneas más arriba, es conveniente orientar la investigación hacia las dos soluciones, puesto que si se trabaja únicamente con la herramienta *Pentaho Business Analytics* los procesos de integración de datos vienen incluidos pero no se

“ANÁLISIS COMPARATIVO DE HERRAMIENTAS DE SOFTWARE LIBRE Y PROPIETARIO PARA LA GESTIÓN DE BIG DATA EN EMPRESAS DE COMERCIALIZACIÓN MASIVA”

pueden apreciar en todo su potencial, cosa que si puede lograrse con la herramienta *Pentaho Data Integration*. Además en esta última se pueden generar los procesos utilizando Big Data con el correspondiente entorno de trabajo.

Para aclarar un poco más la relación con Big Data, es necesario citar el siguiente gráfico por la compañía Pentaho (2015):

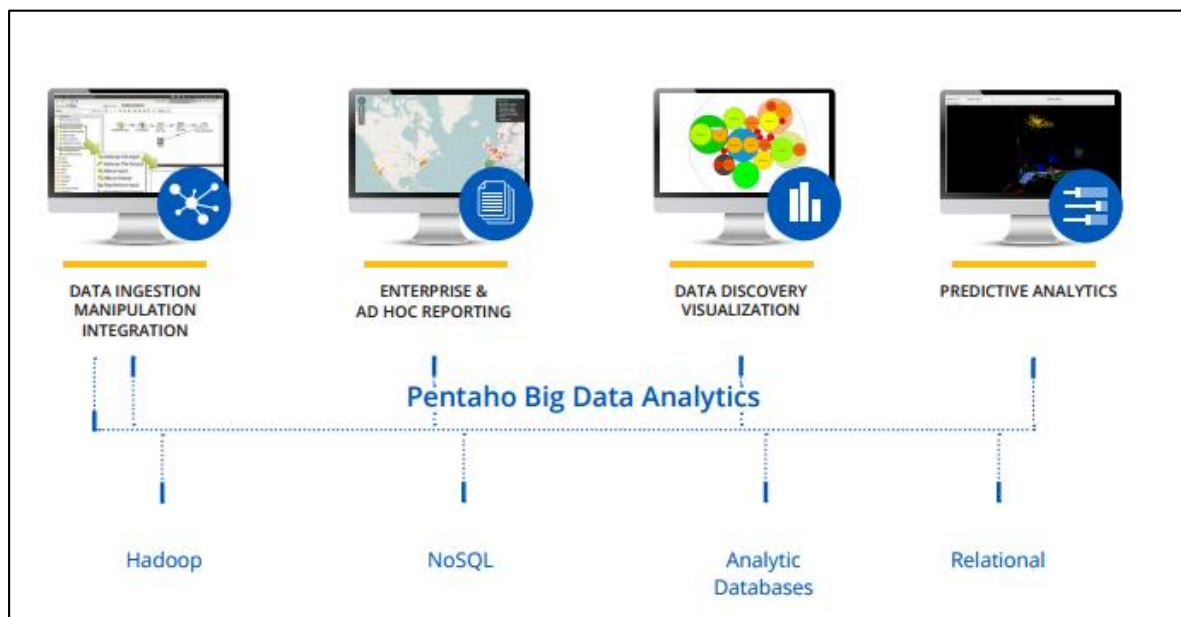


Figura 4-01: Análisis de Big Data y Gestión de Datos usando Pentaho. Fuente: Pentaho (2015)

La interpretación del gráfico anterior supone lo siguiente: Partiendo de cualquiera de las categorías de fuentes relacionadas a Big Data, la herramienta de Pentaho está vinculada directamente a los procesos de manipulación e integración de datos, generación de reportes adecuados según la acción, visualización del descubrimiento de datos y finalmente un análisis predictivo. En otras palabras, todo el enfoque relacionado con Big Data la herramienta de Pentaho lo resume en el gráfico anterior.

Para concluir con este punto se adjuntan algunas ilustraciones que indican un panorama de visión general de las soluciones propuestas anteriormente:

“ANÁLISIS COMPARATIVO DE HERRAMIENTAS DE SOFTWARE LIBRE Y PROPIETARIO PARA LA GESTIÓN DE BIG DATA EN EMPRESAS DE COMERCIALIZACIÓN MASIVA”

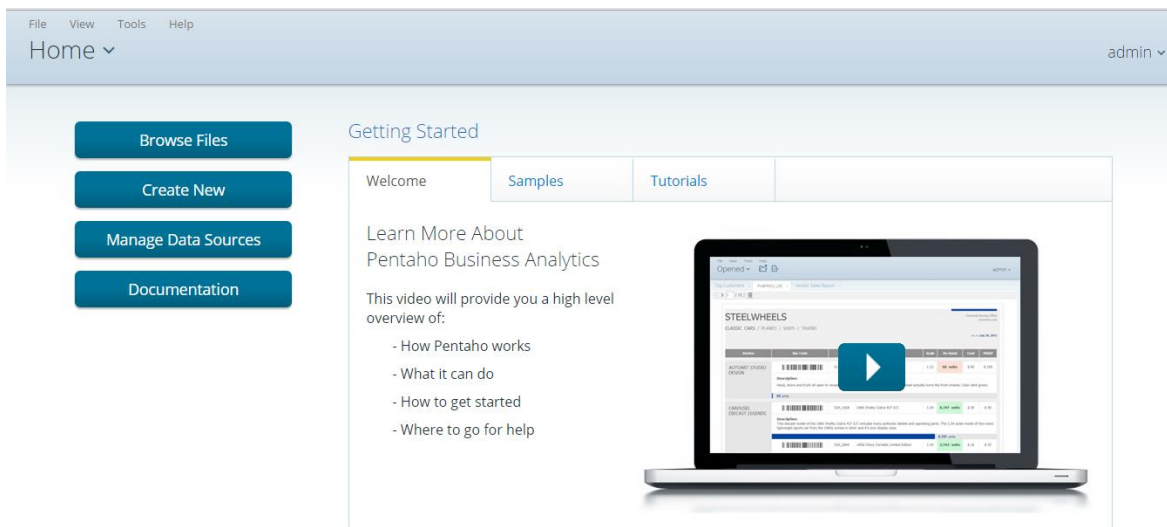


Figura 4-02: Herramienta Pentaho Business Analytics (Pantalla Principal)

File View Tools Help

Opened

Inventory List (report) [ES] x CTools Dashboard x Leading Product Lines (pivot table) x

1 Filter

		Years								
Territory	Line	2003			2004			2005		
		Sales	Quantity	Unit Sales	Sales	Quantity	Unit Sales	Sales	Quantity	Unit Sales
APAC	Classic Cars	\$115.011	1.052	\$109	\$199.372	1.785	\$112	\$97.574	1.015	\$96
	Vintage Cars	\$111.639	1.243	\$90	\$147.212	1.587	\$93	\$105.688	1.067	\$99
	Motorcycles	\$60.789	654	\$93	\$63.159	540	\$117	\$65.870	658	\$100
	Trucks and Buses	\$11.298	91	\$124	\$80.634	801	\$101	\$53.735	488	\$110
	Planes	\$42.663	456	\$94	\$67.681	723	\$94	\$11.082	151	\$73
APAC Total		\$341.400	3.496	\$98	\$558.057	5.436	\$103	\$333.948	3.379	\$99
EMEA	Classic Cars	\$691.273	5.853	\$118	\$1.015.790	8.976	\$113	\$384.538	3.463	\$111
	Vintage Cars	\$263.695	3.094	\$85	\$504.062	5.472	\$92	\$83.324	1.094	\$76
	Motorcycles	\$141.836	1.428	\$99	\$204.042	2.177	\$94	\$161.260	1.501	\$107
	Trucks and Buses	\$228.699	2.261	\$101	\$185.421	1.558	\$119	\$86.859	836	\$104
	Planes	\$154.519	1.723	\$90	\$209.128	2.326	\$90	\$128.008	1.464	\$87
EMEA Total		\$1.480.021	14.359	\$103	\$2.118.443	20.509	\$103	\$843.989	8.358	\$101
Japan	Classic Cars	\$120.696	898	\$134	\$42.071	307	\$137	\$18.835	122	\$154
	Planes	\$60.556	677	\$89	\$49.177	547	\$90	-	-	-
	Trucks and Buses	\$44.498	415	\$107	\$13.349	102	\$131	-	-	-

Figura 4-03: Herramienta Pentaho Business Analytics (Tabla dinámica)

“ANÁLISIS COMPARATIVO DE HERRAMIENTAS DE SOFTWARE LIBRE Y PROPIETARIO PARA LA GESTIÓN DE BIG DATA EN EMPRESAS DE COMERCIALIZACIÓN MASIVA”

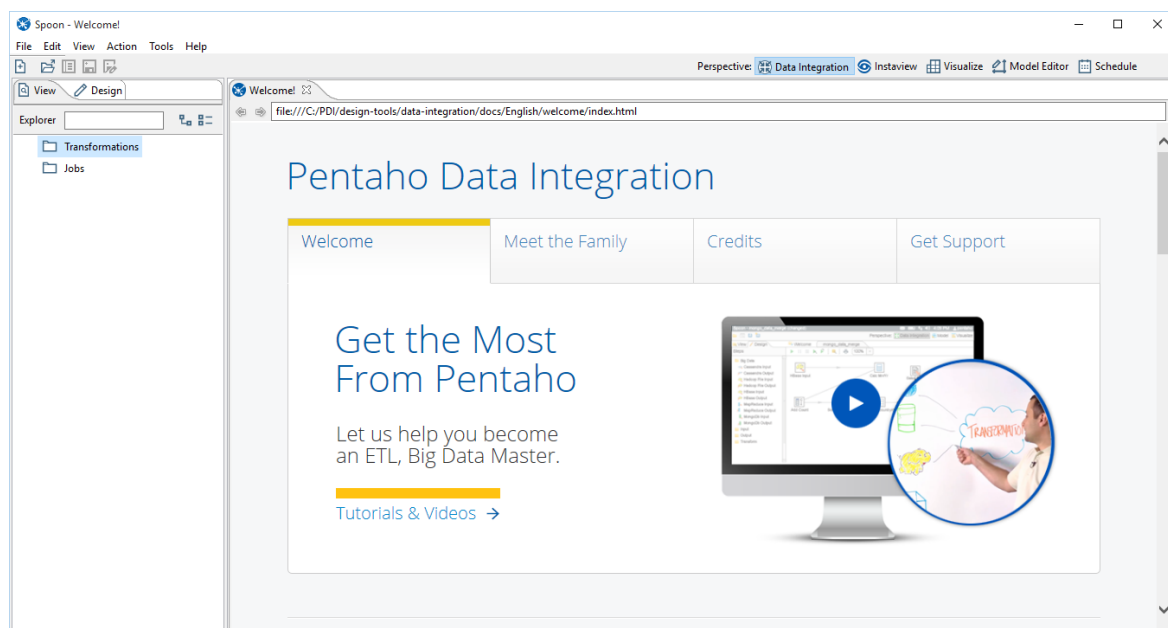


Figura 4-04: Herramienta *Pentaho Data Integration* (Pantalla Principal)

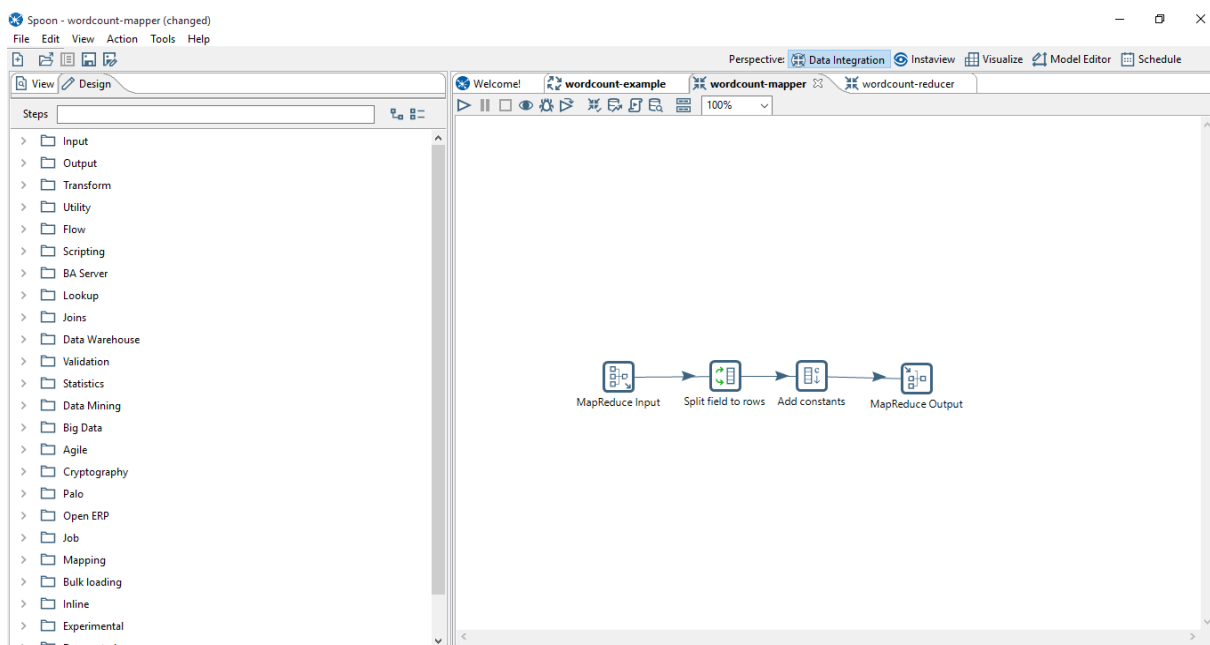


Figura 4-05: Herramienta *Pentaho Data Integration* (Procesos MapReduce)

4.2. Herramientas de Software Propietario

Una vez descrita la herramienta de trabajo que corresponde a software libre, ahora por el contrario se hablará de las herramientas de software propietario. De igual manera me parece conveniente usar la misma temática que en el punto anterior, que justamente comenzaba con preguntarse a uno mismo el significado, dar la apreciación personal y después citar la parte conceptual que incluye las principales características.

En este punto, cuando se habla de software propietario inmediatamente se puede inferir que es lo contrario a un software libre. Mientras el software libre es gratuito en algunas ocasiones, sin restricciones y con libertad para su uso, el software propietario es pagado y con algunas limitaciones en cuanto a sus funcionalidades y libertades para usarlo. Así mismo la anterior es una apreciación o definición en términos generales, por ende vale la pena explicar más detallada y concisamente las características principales del software propietario con el fin de esclarecer su concepto y entender las diferencias puntuales con el software libre.

Para esto voy a hacer referencia a tres fuentes que mencionan lo anterior de un modo claro y preciso:

Culebro Juárez, Gómez Herrera y Torres Sánchez (2006) en su artículo *Software libre vs software propietario: ventajas y desventajas* manifiestan que el software no libre o también llamado software privativo “se refiere a cualquier programa informático en el que los usuarios tienen limitadas las posibilidades de usarlo, modificarlo o redistribuirlo (con o sin modificaciones), cuyo código no esté disponible o el acceso a éste se encuentre restringido”.

Por su parte la *Free Software Foundation* (FSF) también se expresa con respecto al software propietario, e indica que éste “es aquel que no es libre. Su uso, redistribución o modificación está prohibida o requiere una autorización”.

La opinión de Martín Carranza Torres (2004) con respecto al software propietario indica que el software propietario “es de dominio privado, porque una persona tiene la titularidad de los derechos de autor y goza de un derecho exclusivo respecto de su utilización. Le niega a otras personas el acceso al código fuente del software y el derecho a copiarlo o estudiarlo” (p.103).

Después de dichas explicaciones, se tiene más claro el concepto de software propietario y a su vez las principales diferencias con el software libre traducidas en limitaciones para distribuciones, copias y mejoras, así como también el pago de una o varias licencias para su uso según sea el caso.

Siguiendo con la temática, después de haber expuesto la parte conceptual del software propietario, ahora es momento de hablar sobre la otra herramienta considerada el otro objeto de estudio en este capítulo. La herramienta elegida será QlikView, igualmente por su popularidad en el mercado y por la integración tanto de Big Data como Business Intelligence, ya mencionados en capítulos anteriores.

El sitio web oficial de la compañía Qlik¹⁵ manifiesta que “es una compañía líder que ofrece soluciones intuitivas para una visualización de datos y una analítica guiada de auto-servicio. Las soluciones de Qlik son fundamentales para extraer conocimiento de información procedente de fuentes muy variadas, explorando las relaciones ocultas entre los datos, lo cual lleva a unos conocimientos que impulsan las buenas ideas”.

¹⁵ URL: <http://www.qlik.com>

Es lógico pensar que esta herramienta llamada QlikView posee algunos beneficios a nivel global con relación a Big Data. De igual manera la compañía Qlik en su sección *Nuestros Productos* → *Big Data* hace referencia a dichos beneficios que son los siguientes:

- Consolidar datos relevantes procedentes de múltiples fuentes, incluidos grandes repositorios de Big Data.
- Acceso al análisis de Big Data sin hacer programación o modelado de datos.
- Exploración de asociaciones entre datos tradicionales y Big Data.
- Visualización de Big Data con gráficos vistosos, potentes y atractivos.
- Permite una toma de decisiones social con una colaboración en tiempo real.

Dentro de esta perspectiva, QlikView ofrece un enfoque específico hacia Big Data, orientado hacia dos alternativas posibles:

1. **Análisis mediante arquitectura en la memoria de QlikView:** En este contexto se comprimen los datos habituales en un factor de 1 a 10, permitiendo la carga de gran cantidad de datos no comprimidos, los cuales representan millones de filas a ser analizadas con tiempos de respuesta que solo son posibles con arquitecturas en memoria.
2. **Análisis mediante la herramienta QlikView Direct Discovery:** Si no se desea cargar todos los datos en la memoria de QlikView, se puede optar por Direct Discovery que aprovecha los datos en memoria y los que se consultan de manera dinámica desde fuentes externas. Con esto se aporta significado y contexto a Big Data debido a que este enfoque les da a los usuarios la información que necesitan en el momento oportuno, así como también pueden seguir explorando datos libremente, sin notar la diferencia sobre su ubicación ya sea ésta en memoria o en sitio.

Una vez mencionadas las alternativas del enfoque de QlikView hacia Big Data, surge la siguiente pregunta: ¿Cuál de las dos alternativas resulta mejor elegir o cuándo elegir cada una? Pues bien, la compañía Qlik en un *White Paper*¹⁶ titulado ***QlikView and Big Data: have it your way*** presenta un cuadro comparativo que responde de manera precisa a pregunta formulada anteriormente: (p.10).

Tabla 4-02: Cuándo elegir 100 % en memoria vs Direct Discovery

¿Cuándo elegir 100 % en memoria?	¿Cuándo elegir Direct Discovery?
Todos los datos necesarios (pertinentes y contextuales) pueden caber en la memoria del servidor (un comprimido de aproximadamente 200 GB).	Los datos no pueden caber en memoria y el encadenamiento de documentos no es una solución viable debido a los requisitos analíticos.
Los usuarios solamente requieren datos agregados o resumidos por hora o por promedios diarios con un nivel récord de detalle durante un período de tiempo limitado.	Los usuarios requieren acceso a nivel de registro de detalles almacenados en una tabla de hechos de gran tamaño que no caben en memoria.
El rendimiento de las consultas de la fuente externa no es satisfactorio o el número de consultas que se esperan de usuarios concurrentes impactarían negativamente en el repositorio de Big Data.	Con limitaciones en ancho de banda de la red tomaría demasiado tiempo para copiar los datos en bruto a un servidor de QlikView. Las consultas de Direct Discovery devuelven datos agregados, por lo tanto, se requiere menos ancho de banda.

Fuente: *White Paper* titulado ***QlikView and Big Data: have it your way*** (2014) por la compañía Qlik. Traducción al español por John Steven Cabezas Jácome (2015).

¹⁶ Término que se refiere a un informe de autoridad y persuasivo que trata en profundidad sobre un tema específico que presenta un problema y ofrece una solución. Fuente: <http://www.inboundemotion.com/blog/%C2%BFqu%C3%A9-es-un-whitepaper>

Siguiendo con el contexto de Big Data relacionado con la herramienta de QlikView, es necesario entender la perspectiva general del mecanismo de flujo de datos y análisis que ofrece la misma. El siguiente gráfico propuesto en el *White Paper* mencionado anteriormente hace referencia justamente a dicho enfoque: (p.4).

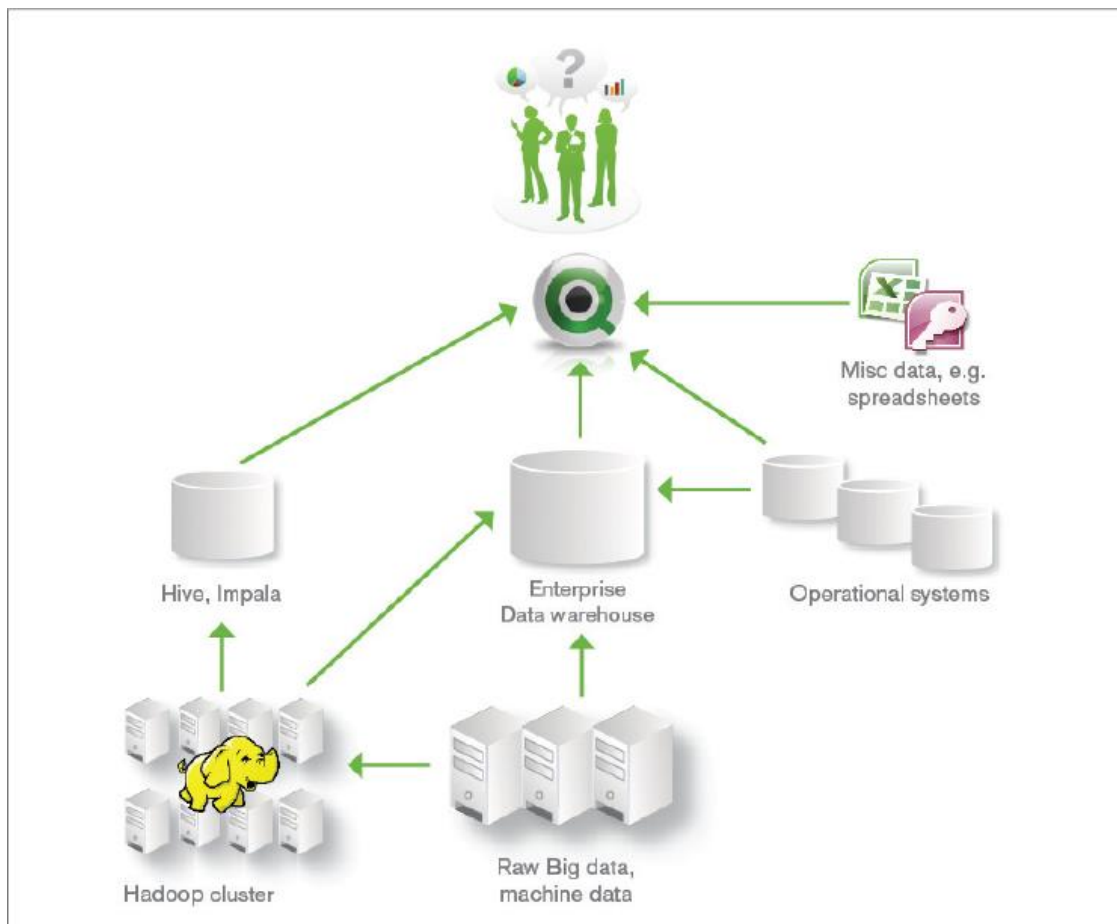


Figura 4-06: Flujo de los datos desde la fuente hasta el análisis. Fuente: *White Paper* titulado *QlikView and Big Data: have it your way* (2014) por la compañía Qlik.

La interpretación del gráfico se sigue así: Se muestra el flujo de datos que comienza desde su parte más simple hasta la más compleja dependiendo por los procesos que pasan los datos donde van adquiriendo mayor valor conforme vayan avanzando de escenario. El clúster de Hadoop permite trabajar con gran cantidad de datos no estructurados para posteriormente procesarlos mediante un Data Warehouse con las correspondientes herramientas de extracción, transformación y carga (ETL), mencionadas en el capítulo 3. Al final se encuentra la etapa de análisis que es donde

“ANÁLISIS COMPARATIVO DE HERRAMIENTAS DE SOFTWARE LIBRE Y PROPIETARIO PARA LA GESTIÓN DE BIG DATA EN EMPRESAS DE COMERCIALIZACIÓN MASIVA”

interviene la herramienta que apoya directamente a las necesidades principales de negocio y a tener un mejor panorama o visión para la toma de decisiones en base a políticas y objetivos estratégicos como ya se conoce.

De igual manera para concluir con este punto es conveniente colocar algunas ilustraciones que indican una visión general de la herramienta de QlikView:

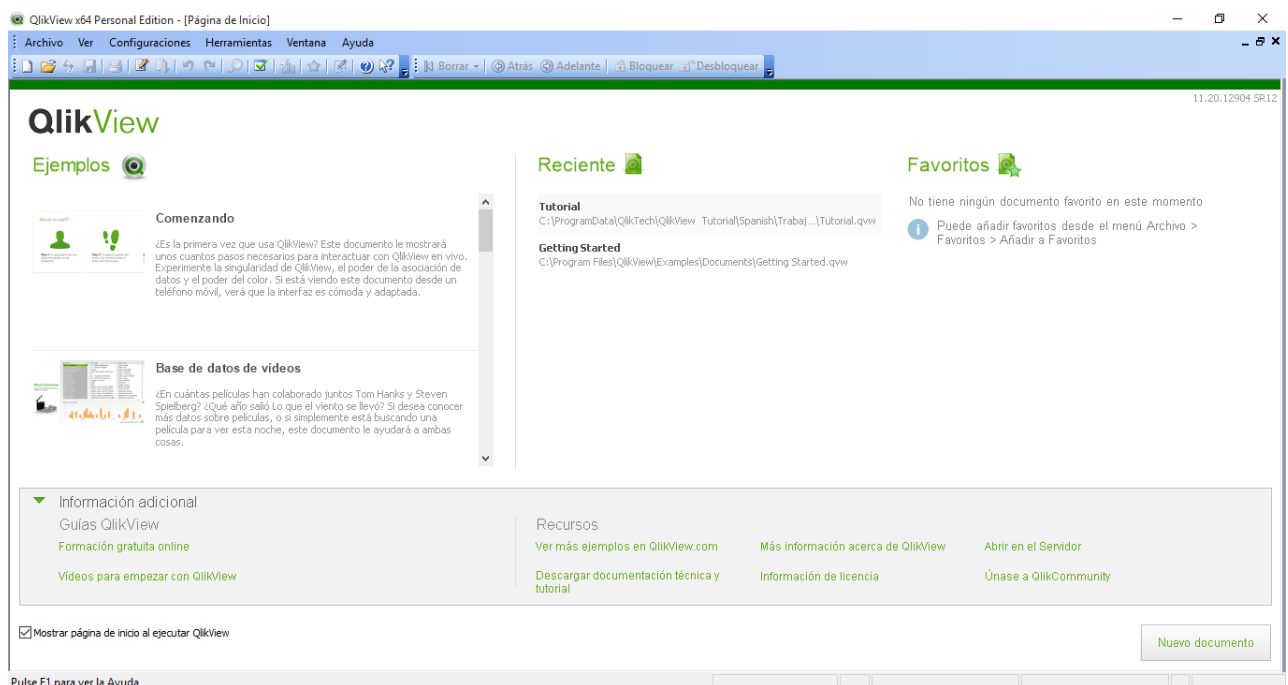


Figura 4-07: Herramienta QlikView Personal Edition (Pantalla Principal)

“ANÁLISIS COMPARATIVO DE HERRAMIENTAS DE SOFTWARE LIBRE Y PROPIETARIO PARA LA GESTIÓN DE BIG DATA EN EMPRESAS DE COMERCIALIZACIÓN MASIVA”

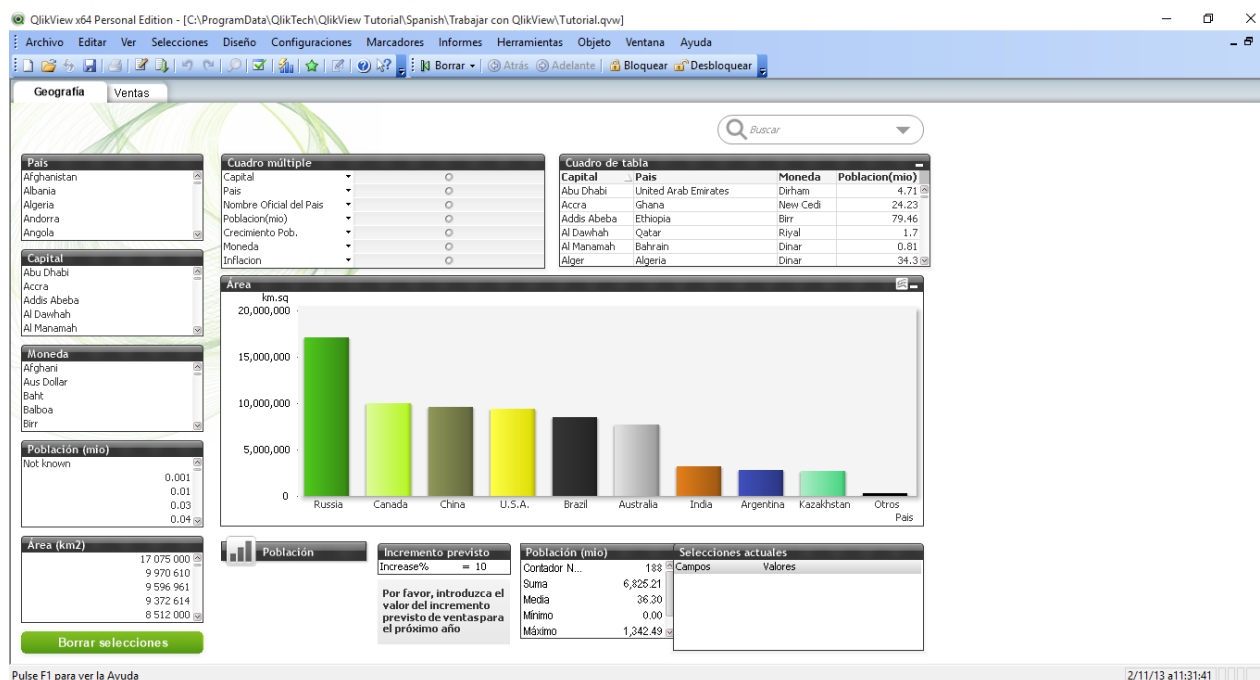


Figura 4-08: Ejemplo de un documento QlikView

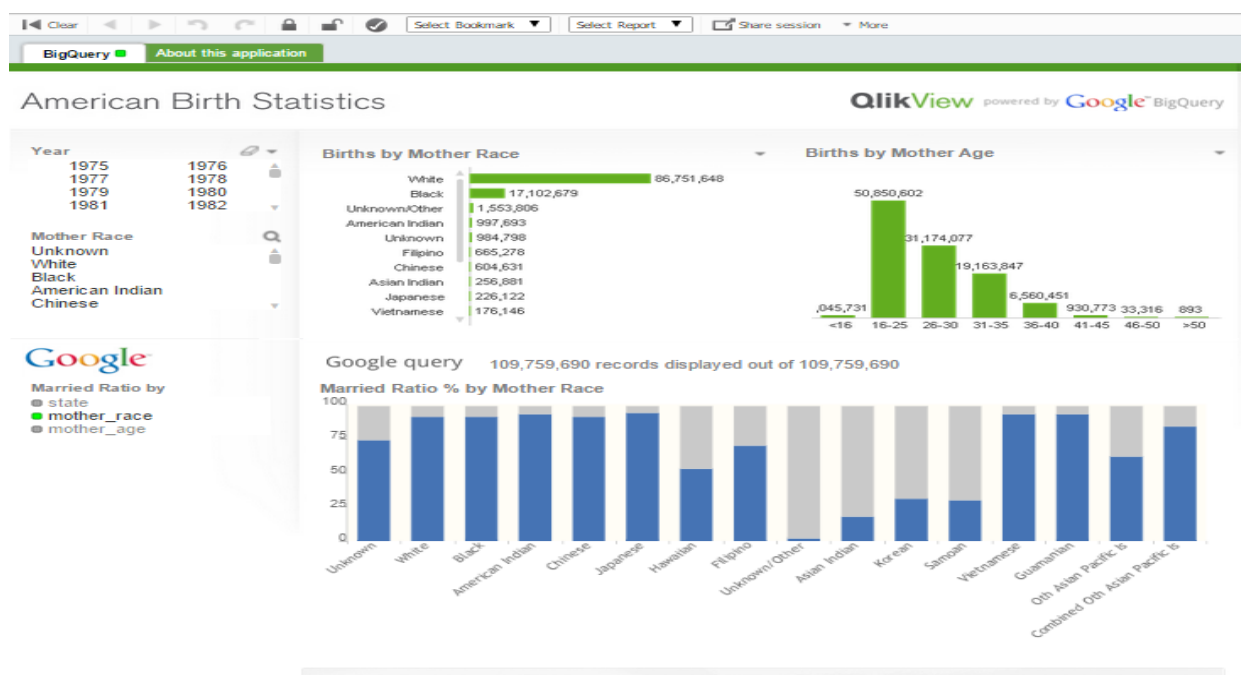


Figura 4-09: Ejemplo de un documento de QlikView con enfoque hacia Big Data usando Google Big Query. Fuente: Qlik Online Demos (2015). URL: <http://us-d.demo.qlik.com/detail.aspx?appName=American%20Birth%20Statistics.qvw>

4.3. Análisis Comparativo

Una vez que se eligió y se habló sobre las herramientas que van a ser objeto de estudio en el presente trabajo, es momento de realizar un análisis comparativo a nivel general entre ambas. La idea es justamente realizar una matriz en donde se describan algunas características que agrupen ventajas y desventajas de las dos herramientas, siendo este el objetivo de la comparación.

Previamente es necesario definir las características o criterios que serán objeto de evaluación en la matriz que correspondan a las dos herramientas. A continuación se citarán los más importantes:

- Licencia de la herramienta: software libre o propietario.
- Oferta de soluciones concretas hacia lo que es Big Data.
- Capacidad de integración de todos sus productos.
- Oferta de servicio técnico como soporte.
- Posibilidad de generación de reportes.
- Costos por licencia.
- Tiempos de implementación.
- Disponibilidad de cubos OLAP para el manejo de información.
- Manejo de interfaces y gráficos.
- Curva de aprendizaje.
- Manejo de vistas para queries y análisis.

Cabe destacar que los criterios mencionados anteriormente son los más sobresalientes para la matriz de análisis comparativo a nivel general. Existen otras características que se explicarán más adelante en los análisis comparativos referentes a funcionalidad, generación de reportes y usabilidad.

Dicho esto, la matriz quedaría de la siguiente manera:

“ANÁLISIS COMPARATIVO DE HERRAMIENTAS DE SOFTWARE LIBRE Y PROPIETARIO PARA LA GESTIÓN DE BIG DATA EN EMPRESAS DE COMERCIALIZACIÓN MASIVA”

Tabla 4-02: Matriz comparativa entre las herramientas de Pentaho Business Analytics y QlikView

CARACTERÍSTICAS	HERRAMIENTAS	
	Pentaho Business Analytics	QlikView
Licencia	Software Libre	Software Propietario
Oferta de soluciones concretas hacia Big Data	Sí, con el producto Pentaho Data Integration (PDI)	Sí, con los dos enfoques: en memoria y Qlik Direct Discovery
Capacidad de integración de todos sus productos	Sí. El PDI viene integrado en la herramienta principal de Pentaho Business Analytics	No está claro
Oferta de servicio técnico como soporte	Solamente a las versiones Enterprise	Sí existe servicio de soporte técnico
Posibilidad de generación de reportes	Sí	Sí
Costos por licencia	La versión Community que es la más utilizada y difundida, no tiene costo	Su licencia comercial tiene costo pero su precio es mucho menor comparado al de otras herramientas propietarias
Tiempos de Implementación	Bastante superior a 3 meses, por lo que el costo de consultoría es mayor	Por lo general dura menos de 3 meses, lo que implica que el costo de consultoría sea menor
Disponibilidad de cubos OLAP para el manejo de información	Sí, aunque esta tecnología tiene limitaciones que acotan el número de dimensiones	No dispone de cubos OLAP, ofrece a cambio una tecnología asociativa
Manejo de interfaces y gráficos	Sí	Sí
Curva de aprendizaje	De 1 a 2 semanas	Inferior a 1 semana
Manejo de vistas para queries y análisis	Sí	Sí

Es importante manifestar que los criterios han sido seleccionados con base a dos puntos de vista: uno publicado en el sitio web Buyto.es¹⁷ con el artículo *Comparativa entre Pentaho y QlikView*, y el otro, el expuesto por Iñigo Castánder, Jon Kepa Gerrikagoitia y Aurkene Alzua-Sorzabal en su artículo *Estudio comparativo de las herramientas de Business Intelligence: Empoderando el criterio de selección a las PYMEs*.

La idea de tomar los elementos de las dos fuentes, mismos que se han colocado como características en la matriz de comparación, es justamente para cubrir y sustentar de mejor manera el análisis comparativo a nivel global que se está planteando. A groso modo se puede pensar que dentro del contexto de Big Data la herramienta *Pentaho Business Analytics* puede ofrecer un mejor enfoque debido a la integración de su módulo PDI (Pentaho Data Integration) visto anteriormente.

Sin embargo, en los siguientes puntos se realizará el análisis comparativo orientado a funcionalidad, generación de reportes y usabilidad, situación que ayudará a esclarecer un poco más esta última afirmación.

4.3.1. Análisis Comparativo Funcional

Está claro que para realizar cualquier tipo de análisis comparativo es necesario establecer criterios, características o métricas que sirvan como objeto de comparación para poder debatir y sacar conclusiones al respecto, tal y como se hizo en el análisis comparativo general.

En este punto se hablará del análisis comparativo funcional. Siguiendo con la misma temática, es necesario plantear un criterio que se convertirá en elemento clave para la comparación de las dos herramientas en cuestión: Pentaho Business Analytics y QlikView.

¹⁷ Buyto.es – portal web de consultores expertos en Business Intelligence desarrollando soluciones en Madrid y resto de España.

Y dicho criterio tiene que ver obviamente con la funcionalidad de cada herramienta, por lo tanto es una buena idea formular la siguiente pregunta: ¿Cuál de las dos herramientas cubre o se adapta mejor a las funciones principales de Business Intelligence? Dicha pregunta será entonces el criterio de comparación en esta parte.

Para poder dar respuesta a la misma, es oportuno mencionar algunos factores indispensables a los cuales las herramientas deben orientarse para considerarlas completamente funcionales. La investigadora Gartner¹⁸ (2015) propone un reporte que contiene un análisis de todas las herramientas orientadas a Business Intelligence en base a 13 capacidades críticas las cuales son:

1. **Trituración de datos del usuario del negocio:** Esta parte incluye actividades como arrastrar y soltar, combinación de datos desde distintas fuentes, creación de modelos analíticos con base en datos múltiples estructurados y métricas definidas por el usuario.
2. **Plataforma interna de integración:** Aquí se incluyen elementos como identidad visual única (look and feel), motor de consulta, metadatos compartidos y todo lo que tenga que ver con los componentes de la plataforma.
3. **Administración de la plataforma de BI:** Capacidades que permiten administrar usuarios, optimizando el rendimiento y la garantía de alta disponibilidad junto con la recuperación ante desastres. Dichas capacidades deben ser comunes a todos los componentes de la plataforma.
4. **Gestión de Metadatos:** Se debe proporcionar una manera robusta y centralizada de búsqueda, captura, almacenamiento, reutilización y publicación de objetos de metadatos, tales como dimensiones, jerarquías, medidas de rendimiento, parámetros y registros.
5. **Despliegue en la nube:** Tiene que ver con la plataforma como servicio y aplicación analítica con capacidades similares para la construcción, despliegue y gestión de datos junto con su análisis respectivo.

¹⁸ Ver cita No.2

6. **Desarrollo e Integración:** La plataforma debe proporcionar un conjunto de herramientas programáticas y visuales, así como también un entorno de trabajo que permita la creación de informes, cuadros de mando, consultas y análisis de contenidos y aplicaciones. Se debe incluir la capacidad de integrar y personalizar componentes de la plataforma en un proceso de negocio determinado.
7. **Exploración interactiva de forma libre:** Permite la exploración de datos a través de la manipulación de imágenes y gráficos con diferentes colores, tamaños, la forma y el movimiento de los objetos visuales que representan los aspectos del conjunto de datos que están siendo analizados. Esto incluye todo tipo de gráficos, mapas geográficos, diagramas de dispersión y otros elementos visuales para fines especiales. Dichas herramientas permiten al usuario analizar los datos con distintas representaciones visuales.
8. **Tableros analíticos y contenidos:** La capacidad de crear cuadros de mando interactivos y contenidos con exploración visual analítica avanzada y análisis geoespacial que pueda ser consumido por otros.
9. **Desarrollo de informes y cuadros de mando:** Proporciona la capacidad de crear informes interactivos con o sin parámetros, añadiendo componentes de visualización tales como medidores, deslizadores o casillas de verificación. Dichos componentes son también utilizados en la creación de los cuadros de mando que pueden representar información operativa y estratégica.
10. **Estilos tradicionales de análisis:** Las consultas ad hoc permiten a los usuarios hacer sus propias preguntas acerca de los datos, sin depender del departamento de TI para crear un informe. Las herramientas deben contener una capa semántica reutilizable para permitir a los usuarios navegar por las fuentes disponibles de datos, métricas predefinidas y jerarquías. El procesamiento analítico en línea (OLAP) permite un estilo de análisis conocido como reordenamiento, que consiste en analizar los datos con consultas rápidas y rendimiento de cálculo. Aquí intervienen las arquitecturas de datos y arquitecturas de almacenamiento.

- 11. Móvil:** Esto tiene que ver con la entrega y desarrollo de contenido para dispositivos móviles aprovechando sus capacidades nativas tales como cámara, táctil, conocimiento de ubicación y de consulta.
- 12. Colaboración e Integración Social:** Permite a los usuarios compartir y discutir la información a través de foros de discusión o chat para de esta manera resolver alguna inquietud puntual con respecto al tema.
- 13. BI embebido:** Capacidades de desarrollo de software con APIs y soporte para estándares abiertos que permitan la creación y modificación de contenidos analíticos, visuales y aplicaciones, así como también su incorporación a un proceso de negocio o hacia una aplicación.

El reporte mencionado fue realizado con base al cuadrante mágico para plataformas de Business Intelligence y analíticas, el mismo que se muestra en la siguiente imagen:



Figura 4-10: Cuadrante Mágico para plataformas de Business Intelligence y analíticas. Fuente: Investigadora Gartner (2015)

De este modo las herramientas que son objeto de estudio en el presente trabajo, tales como son Pentaho Business Analytics y QlikView, se encuentran en el primer y tercer cuadrante respectivamente. Se puede apreciar que Pentaho se encuentra en el cuadrante de los *Niche Players* mientras que QlikView es una herramienta posicionada en el cuadrante de los *Leaders*. A continuación se dará una breve explicación con respecto a ambos términos.

La misma investigadora Gartner en un artículo de investigación titulado ***Gartner Magic Cuadrant***, que traducido al español sería *Cuadrante Mágico de Gartner*, explica muy brevemente dichos términos de la siguiente manera:

- **Niche Players:** Plataformas que se enfocan en un pequeño segmento o que se encuentran ligeramente orientadas a innovación o superación hacia las demás.
- **Leaders:** Plataformas que se enfocan en su visión actual y están en buena posición para el futuro.

Continuando con el análisis propuesto por la investigadora Gartner en relación con el cuadrante mágico, se citarán algunos puntos fuertes y débiles de las dos herramientas con relación a las 13 capacidades críticas que ya se habían visto:

PENTAHO BUSINESS ANALYTICS

Puntos Fuertes:

- ✓ BI embebido.
- ✓ Estilos tradicionales de análisis.
- ✓ Tableros analíticos y contenidos.
- ✓ Desarrollo e Integración.
- ✓ Plataforma interna de integración.
- ✓ Administración de la plataforma de BI.
- ✓ Gestión de Metadatos.

Puntos Débiles:

- ☒ Colaboración e integración social.
- ☒ Móvil.
- ☒ Despliegue en la nube.

QLIKVIEW

Puntos Fuertes:

- ✓ Estilos tradicionales de análisis (memoria).
- ✓ Exploración interactiva de forma libre.
- ✓ Móvil.
- ✓ Colaboración e integración social.
- ✓ Tableros analíticos y contenidos.

Puntos Débiles:

- ☒ Despliegue en la nube.
- ☒ Gestión de Metadatos.
- ☒ Plataforma interna de integración.
- ☒ BI embebido.

Para concluir con este punto, merece la pena mencionar nuevamente al sitio web Buyto.es¹⁹ puesto que allí se proponen algunas cuestiones que a mi parecer encajan perfectamente en el ámbito funcional. La herramienta de Pentaho Business Analytics resultaría ser la que mejor se orienta funcionalmente puesto que posee un modelo multidimensional centralizado con aproximación a metadatos, así como también posee su propio módulo de integración de datos que contiene también las herramientas de transformación, extracción y carga (ETL).

¹⁹ Ver cita No.17

Por el contrario, QlikView utiliza la lógica asociativa, la que consiste en realizar análisis, cálculos y consultas directamente en memoria con tiempos de respuesta bastante decentes. Aunque es muy útil para la elaboración de prototipos sencillos enfocados a cuadros de mando sencillos también, se pierde eficiencia cuando se quiere pasar a un sistema BI más complejo con análisis OLAP e informes así mismo más sofisticados. Es por esta razón que carece que metadatos centralizados, así como también de una suite ETL para integración de datos.

4.3.2. *Análisis Comparativo en Generación de Reportes*

Es este punto resulta un poco obvio pensar cual será el criterio de comparación entre las dos herramientas. Dicho criterio va orientado hacia la generación de reportes de cada una, siendo la pregunta principal: ¿Cuál de las dos herramientas puede generar reportes que cubran de mejor manera cierto tipo de necesidades? A mi parecer no es suficiente con solo responder a esta pregunta y tomarla como único criterio para la comparación, porque rápidamente se podría inferir que la herramienta de *Pentaho Business Analytics* posee un mejor enfoque y orientación cuando se trata de cubrir necesidades orientadas a Big Data.

Pero la idea en este punto es a partir del criterio general previamente definido, mencionar algunos criterios secundarios que puedan apoyar a la comparación de una mejor manera. Dichos criterios secundarios son los siguientes:

- Esquemas de selección e interacción en una sola vista.
- Análisis de contenido con metadatos.
- Manejo de grandes volúmenes de información.
- Tiempos de respuesta.
- Manejo de asociaciones de datos.

Para lo cual es necesario realizar un pequeño cuadro comparativo con una breve explicación de los criterios antes mencionados:

“ANÁLISIS COMPARATIVO DE HERRAMIENTAS DE SOFTWARE LIBRE Y PROPIETARIO PARA LA GESTIÓN DE BIG DATA EN EMPRESAS DE COMERCIALIZACIÓN MASIVA”

Tabla 4-03: Cuadro comparativo con criterios relacionados a la generación de reportes. Pentaho Business Analytics vs QlikView

CRITERIOS DE COMPARACIÓN	HERRAMIENTAS	
	Pentaho Business Analytics	QlikView
Esquemas de selección e interacción en una sola vista	No. El reporte solamente se genera por selección de campos	Sí. Este tipo de esquema permite interacción entre la información que se desea conocer en base a un criterio
Análisis de contenido con metadatos	Sí, mediante la creación de un modelo de metadatos previo	No, debido a la carencia de metadatos centralizados
Manejo de grandes volúmenes de información	Sí, debido a la disponibilidad de cubos OLAP	No se consiguen reportes a gran nivel de detalle para grandes listados
Tiempos de respuesta	Decentes pero algunas veces no tan rápidos como en QlikView	Más rápidos debido al manejo en memoria
Manejo de asociaciones de datos	Manualmente	Automáticamente

Como conclusión para este punto, se puede inferir que la herramienta de QlikView se ajusta mejor a la característica de la generación de reportes, debido a que posee un esquema de interacción más completo que permite analizar información relacionada con respecto a una variable en diferentes campos, además que brinda un tiempo de respuesta más rápido en comparación con la herramienta de Pentaho Business Analytics.

Pero a esta última no hay que quitarle crédito si lo que se desea es generar reportes con base a metadatos y gran cantidad de información. A continuación se colocarán unas capturas de pantalla con el fin de ilustrar ambos casos en lo que se refiere a generación de reportes:

“ANÁLISIS COMPARATIVO DE HERRAMIENTAS DE SOFTWARE LIBRE Y PROPIETARIO PARA LA GESTIÓN DE BIG DATA EN EMPRESAS DE COMERCIALIZACIÓN MASIVA”

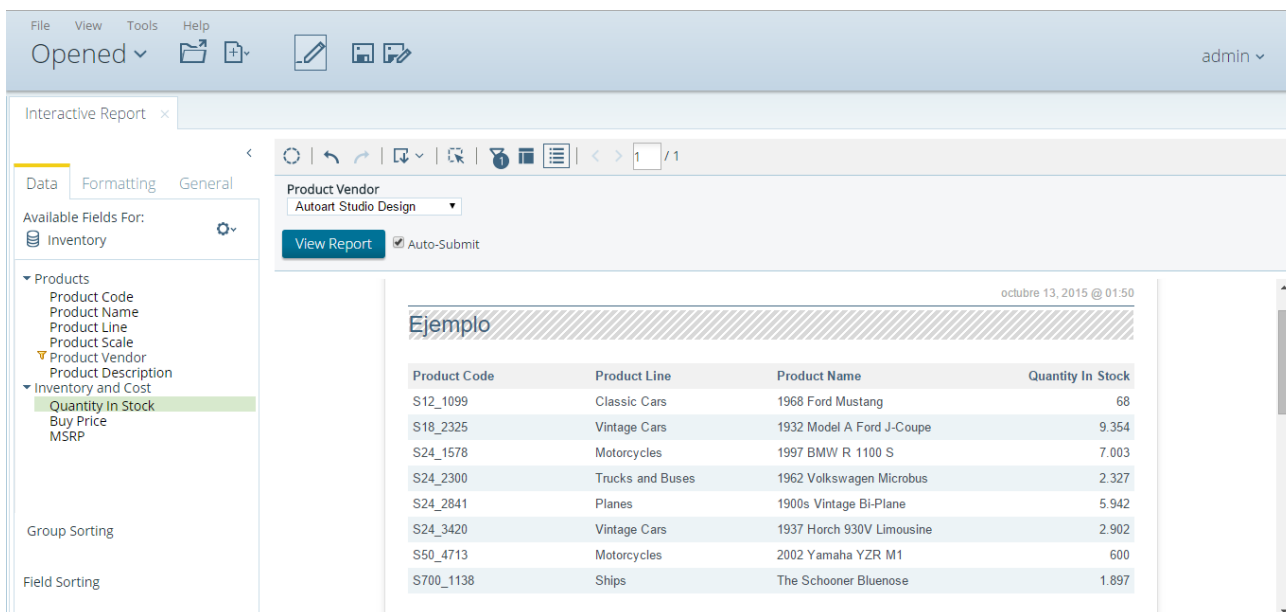


Figura 4-11: Ejemplo de un reporte generado en la herramienta de Pentaho Business Analytics.

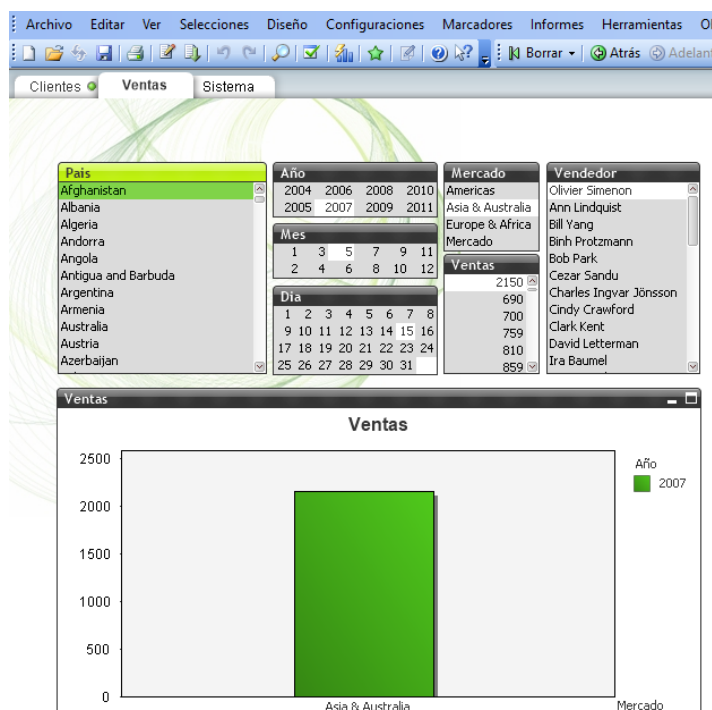


Figura 4-12: Ejemplo de un reporte generado en la herramienta de QlikView

4.3.3. *Análisis Comparativo en Usabilidad*

Ha llegado el momento de realizar el análisis comparativo en cuanto a usabilidad. Antes de definir alguno o varios criterios de comparación en este punto, considero conveniente mencionar el significado del término usabilidad para evitar posibles confusiones. Se puede relacionar la palabra usabilidad con facilidad de manejo, es decir algo que no resulte muy complejo para los usuarios, por lo que se puede inferir que la usabilidad es la característica que indica facilidad de manejo dentro de un entorno que resulta cómodo y satisfecho para los usuarios.

Y es justamente con respecto a esta definición que se deriva el criterio o la pregunta principal para este análisis. Lógicamente la pregunta más importante será: ¿Cuál de las dos herramientas es más fácil de usar para los usuarios? De igual manera que en el capítulo anterior es oportuno definir criterios secundarios que apoyen a la comparación a realizar. Dichos criterios secundarios son los siguientes:

- Curva de aprendizaje.
- Apariencia estética.
- Nivel de complejidad en la manipulación de datos.
- Compatibilidad con distintos sistemas operativos.
- Flexibilidad.
- Tiempos de respuesta a nivel de tareas.

Los criterios de curva de aprendizaje y de tiempos de respuesta ya se comentaron en los puntos anteriores. La curva de aprendizaje encaja muy bien en esta comparación puesto que mientras sea menor el tiempo de aprendizaje, se puede pensar que brinda mayor facilidad de adaptación e integración a las funcionalidades que ofrece la herramienta. Los tiempos de respuesta ahora ya abarcan un entorno más global en cuanto a ejecución de tareas generalizadas. Una vez aclarados tales criterios, el cuadro comparativo quedaría de la siguiente manera:

“ANÁLISIS COMPARATIVO DE HERRAMIENTAS DE SOFTWARE LIBRE Y PROPIETARIO PARA LA GESTIÓN DE BIG DATA EN EMPRESAS DE COMERCIALIZACIÓN MASIVA”

Tabla 4-04: Cuadro comparativo con criterios relacionados a la usabilidad. Pentaho Business Analytics vs QlikView

CRITERIOS DE COMPARACIÓN	HERRAMIENTAS	
	Pentaho Business Analytics	QlikView
Curva de aprendizaje	De 1 a 2 semanas	Inferior a 1 semana
Apariencia Estética	Interfaces con gráficos intuitivos y completos, esquema estático y dinámico	Interfaces con gráficos intuitivos y completos, esquema dinámico de selección de datos
Nivel de complejidad en la manipulación de datos	Alto y robusto dependiendo de las fuentes de datos	Menor, debido a un enfoque de cuadro de mando con un modelo de datos sencillo
Compatibilidad con distintos sistemas operativos	Sí. Windows, Linux y Mac OS ²⁰	²¹ Muy dependiente de arquitecturas y entornos Windows
Flexibilidad	Sí	Sí
Tiempos de respuesta a nivel de tareas	Aceptable. Se realiza un manejo relacional	Más rápidos debido al manejo en memoria

En conclusión se puede inferir que la herramienta de QlikView es más sencilla de usar en comparación con la herramienta de Pentaho Business Analytics, debido a que la curva de aprendizaje es de un tiempo menor, no se tiene mucho nivel de complejidad en modelos de datos para la manipulación, posee un esquema dinámico de selección de datos y tiempos de respuesta más rápidos a nivel de tareas.

Aunque QlikView sea muy dependiente de entornos Windows, podría considerarse más usable siempre y cuando la cultura empresarial esté ajustada ha dicho sistema operativo.

²⁰ Aporte realizado por Santiago Roldán Zuluaga (2013), consultor empresarial en temas de mercadeo por Internet, asesor e implementador de plataformas de comercio electrónico. Fuente: <http://www.santiagoroldan.com/aplicaciones-de-codigo-abierto-para-entornos-empresariales/>

²¹ Fuente: Qlik(2014)

“ANÁLISIS COMPARATIVO DE HERRAMIENTAS DE SOFTWARE LIBRE Y PROPIETARIO PARA LA GESTIÓN DE BIG DATA EN EMPRESAS DE COMERCIALIZACIÓN MASIVA”

Finalmente, se presenta el siguiente cuadro comparativo de resumen que contiene los criterios más importantes en base a todo el análisis realizado con sus respectivos pesos que representan su importancia dentro de cada categoría:

Tabla 4-05: Cuadro de resumen con los criterios más importantes y sus respectivos pesos. Pentaho Business Analytics vs QlikView

CRITERIOS DE COMPARACIÓN	HERRAMIENTAS		
	Puntaje	Pentaho Business Analytics	QlikView
<u>USABILIDAD</u>	30	20	24
Curva de aprendizaje	3	1	3
Apariencia Estética	3	2	3
Nivel de complejidad en la manipulación de datos	6	5	3
Compatibilidad con distintos sistemas operativos	5	4	2
Flexibilidad	4	3	4
Popularidad y uso en las empresas	6	3	6
Tiempos de respuesta a nivel de tareas	3	2	3
<u>GENERACIÓN DE REPORTE</u>	30	21	22
Esquemas de selección e interacción en una sola vista	7	3	7
Análisis de contenido con metadatos	5	5	1
Manejo de grandes volúmenes de información	7	7	3
Tiempos de respuesta	5	3	5
Manejo de asociaciones de datos	6	3	6

“ANÁLISIS COMPARATIVO DE HERRAMIENTAS DE SOFTWARE LIBRE Y PROPIETARIO PARA LA GESTIÓN DE BIG DATA EN EMPRESAS DE COMERCIALIZACIÓN MASIVA”

<u>FUNCIONALIDAD</u>	40	32	19
BI Embebido	6	6	1
Estilos tradicionales de análisis	4	3	4
Tableros analíticos y contenidos	4	4	4
Plataforma interna de integración	8	8	1
Gestión de Metadatos	7	7	1
Colaboración e Integración social	4	2	4
Móvil	3	1	3
Despliegue en la nube	4	1	1
TOTAL	100	73	65

En el cuadro anterior se puede apreciar que la categoría de la funcionalidad representa un 40 %, mientras que la generación de reportes junto con la usabilidad representa el 30 %. Por lo analizado en la comparación funcional pienso que allí es donde se tiene criterios de mayor peso para el análisis puesto que explican el mecanismo de funcionamiento de las herramientas adaptado a la realidad de las empresas. La usabilidad y la generación de reportes son aspectos que se destacan en menor magnitud, es decir que son complementarios a la funcionalidad para poder realizar un análisis de manera general.

Se introdujo un nuevo criterio denominado “Popularidad y uso en las empresas” con el fin de mostrar la herramienta que es más utilizada en el medio empresarial. Curiosamente la herramienta de QlikView es la que más puntaje se lleva y una de las más usadas actualmente debido a su posición como herramienta líder en el cuadrante mágico de Gartner del año 2015, mencionado anteriormente como la Figura 4-10.

4.4. Recomendaciones técnicas para empresas de comercialización masiva

En este punto se definirá un manual de buenas prácticas a manera de algoritmo para la migración o adaptación entre plataformas de Business Intelligence con orientación a Big Data, independientemente si son de software libre o propietario. Dicho esto el algoritmo será el siguiente:

1. Definir métricas e indicadores de productividad (KPI) claros y objetivos que permitan medir eficientemente la gestión del área operativa a analizar.
2. Iniciar el proceso de extracción, transformación y carga (ETL).
 - a. Generar consultas sumariadas de la información que servirán como insumo para los indicadores de producción.
3. Diseñar el data warehouse (DWH)
 - a. Crear un repositorio histórico de variables consultadas e indicadores calculados.
4. Estructurar la información de manera jerárquica para mostrar en el reporte.
 - a. Definir criterios importantes que se distribuirán estratégicamente en el reporte gráfico.
5. Diseñar el reporte considerando limitaciones técnicas y de automatización dependiendo del software utilizado

Cada uno de los puntos anteriores se relacionan con el gráfico propuesto y explicado en el apartado 3.2 Figura 3-02 referente a la arquitectura de una solución de Business Intelligence, y es lógico dado que el sentido o la idea de estos 5 pequeños puntos no es otra que orientar a los analistas de datos hacia un enfoque global de procedimiento tanto en herramientas de software libre como en herramientas de software propietario.

También vale la pena manifestar que este enfoque es completamente aplicable en cualquiera de las áreas de funcionamiento en las empresas de comercialización masiva, situación expuesta y analizada en el apartado 2.2.3.

5. CAPÍTULO 5: CONCLUSIONES, RECOMENDACIONES Y BIBLIOGRAFÍA

5.1. Conclusiones

1. El análisis comparativo entre herramientas de software libre y propietario para la gestión de Big Data en empresas de comercialización masiva es posible en la medida que se tenga pleno conocimiento del giro del negocio, incluyendo políticas y objetivos estratégicos como los pilares fundamentales para el análisis en las distintas áreas de operación.
2. Después de haber realizado la investigación del funcionamiento de las empresas de comercialización masiva, se logró establecer la relación entre las áreas más importantes para operar una empresa junto con los elementos principales de la comercialización masiva, situación explicada en el apartado 2.2.3 Figura 2-04.
3. Se investigó y se logró definir un manual de buenas prácticas que permite migrar a cualquier analista de información entre herramientas de software libre y propietario. Esta situación se propuso como recomendaciones técnicas para las empresas de comercialización masiva dentro del apartado 4.4.
4. De acuerdo al presente trabajo en todo el capítulo 4 se pueden evidenciar conceptos de software libre y propietario como base para la comparación de las herramientas con sus ventajas y desventajas más sobresalientes dentro del contexto de Big Data, utilizando criterios principales y secundarios como metodología de análisis.
5. El Business Intelligence tiene una relación directa con la generación de conocimiento, puesto que es un elemento fundamental y necesario para la toma de decisiones dentro de una organización, donde la información ha adquirido cierto nivel de valor para ser analizada.

6. Según lo expuesto en el apartado 4.3, la herramienta de software libre analizada Pentaho Business Analytics se adapta mejor al contexto de las empresas de comercialización masiva particularmente dentro del enfoque de Big Data debido a que posee un propio módulo de integración de datos que permite dar un mejor apoyo y soporte a dicho enfoque en lo que se refiere a procesos, análisis y manipulación de diversas fuentes de datos.
7. Se puede pensar que con la temática del software libre es posible generar fuentes de empleo a nivel técnico, pero con la desventaja de que se puede generar mucha dependencia de uno o varios expertos en el tema.
8. El uso de cubos de información permite explorar la misma de mejor manera y con la intención de realizar un análisis más profundo y robusto.
9. La implementación de cada reporte debería planearse como un proyecto independiente donde se manejen tiempos y recursos de manera eficiente.
10. Para el diseño de indicadores se deben definir sus condiciones de funcionamiento de tal manera que el dominio de la función que gobierna el indicador represente objetivamente el fenómeno.
11. Es importante crear sistemas cuantitativos de medición de efectividad en los reportes así como también del área que los genera con la intención de llegar a procesos de calidad y mejora continua.

5.2. Recomendaciones

1. Realizar un caso de estudio en algunas empresas con el fin de cuantificar el cumplimiento de los objetivos y funcionalidades de las herramientas de Business Intelligence, junto con la definición de parámetros de comparación más específicos y con mayor nivel de detalle.
2. Plantear la posibilidad de generar listados de evidencia a partir de los reportes gráficos generados por las herramientas dentro del contexto de análisis de datos. Esta actividad puede ser una fuente potencial para mejorar la cultura de toma de decisiones en las organizaciones.
3. Sería bueno realizar un estudio más detallado en cada una de las áreas operativas en las empresas de comercialización masiva con el fin de explotar las

potencialidades y objetivos de cada una apoyados en las herramientas de Business Intelligence.

4. Si bien Big Data sigue siendo el boom en análisis de datos e información, ahora existe Data Science que abarca un campo de análisis más amplio, en donde se incluyen más criterios de análisis, profundización, confiabilidad y veracidad acerca de los datos. No sería descabellado realizar investigaciones futuras dentro de esta temática.
5. Para futuros análisis se deberían considerar criterios más estandarizados, más profundos y ampliamente aceptados por distintas entidades líderes en el desarrollo de software y análisis de datos. Con dichos estándares el análisis puede resultar más efectivo y preciso.
6. Es importante conocer toda la variedad de técnicas y herramientas que trabajan bajo el concepto de Business Intelligence y sobre todo entender las posibles maneras en las que pueden beneficiar a una organización en particular. Aquí también vale la pena considerar su capacidad de combinación con otras tecnologías que puedan adaptarse a un enfoque determinado.
7. Es completamente válido para futuras ocasiones probar las versiones pagadas del software propietario con el fin de explotar todos sus módulos y funcionalidades en procesos de comparación con otras herramientas.
8. Ajustar el contexto y la realidad actual de las empresas hacia sus políticas y objetivos estratégicos como punto de partida para poder adquirir una cultura de toma de decisiones, característica fundamental para mejorar el rendimiento y la productividad dentro de las mismas.
9. Analizar detenidamente requerimientos mínimos del sistema y alternativas de soporte técnico antes de implantar cualquier solución de Business Intelligence, ya que no todas las herramientas funcionan de la misma manera en entornos similares.
10. Realizar estudios similares con otras herramientas tomando como base los criterios más importantes expuestos a manera de resumen en la Tabla 4-05. Cabe recalcar que durante el análisis los valores de los pesos o porcentajes dependerán exclusivamente del caso de estudio y de las herramientas que se utilicen para el mismo.

11. Definir métricas, indicadores y objetivos lo suficientemente claros para el área operativa donde se quiera llevar a cabo un análisis de datos e información.
12. Considerar el proceso de extracción, transformación y carga como factor clave para la construcción, evolución y desenvolvimiento del data warehouse haciendo uso de la arquitectura de metadatos.
13. Tener en cuenta las limitaciones de la herramienta utilizada para procesos de extracción, transformación y carga como también en el diseño de un data warehouse, con el objetivo de sacar el mayor provecho que ofrece dicha herramienta.
14. Establecer criterios de diseño con base a la información estructurada previamente para poder realizar los reportes correspondientes y posteriormente analizar los resultados obtenidos.

5.3. Bibliografía

- ✓ BARRANCO FRAGOSO, Ricardo. *¿Qué es Big Data?* Internet. <https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/>. Acceso: (18 de julio de 2015).
- ✓ BuyTo (Sitio Web). *Comparativa entre Pentaho y QlikView*. Internet. <http://www.buyto.es/general-business-intelligence/comparativa-y-diferencias-entre-pentaho-y-qlikview>. Acceso: (6 de octubre de 2015).
- ✓ CAFERRI, Cecilia. *7 áreas para operar una empresa*. Internet. <http://negocios.about.com/od/Administracion/tp/7-Areas-Para-Operar-Una-Empresa.htm>. Acceso: (27 de julio de 2015).
- ✓ CURTO DÍAZ, Josep. *Introducción al Business Intelligence*. Editorial UOC, 2ª edición, 2012.
- ✓ Gartner. *Gartner IT Glossary: Big Data*. Internet. <http://www.gartner.com/it-glossary/big-data>. Acceso: (16 de julio de 2015).

“ANÁLISIS COMPARATIVO DE HERRAMIENTAS DE SOFTWARE LIBRE Y
PROPIETARIO PARA LA GESTIÓN DE BIG DATA EN EMPRESAS DE
COMERCIALIZACIÓN MASIVA”

- ✓ Gartner. *Gartner IT Glossary: Business Intelligence (BI)*. Internet. <http://www.gartner.com/it-glossary/business-intelligence-bi/>. Acceso: (12 de agosto de 2015).
- ✓ Gartner. *Magic Cuadrant for Business Intelligence and Analytics Platforms*. Internet. <http://www.gartner.com/technology/reprints.do?id=1-2ADAAAYM&ct=150223&st=sb>. Acceso: (8 de octubre de 2015).
- ✓ GNU. *¿Qué es GNU?* Internet. <http://www.gnu.org/>. Acceso: (5 de octubre de 2015).
- ✓ GRAY, Douglas A. & CYR, Donald G. *Marketing de productos: Guía de planificación para pequeñas empresas*. Barcelona, Editorial Granica S.A, 3ª edición, 1993.
- ✓ IBM. “Big Data: What is Big Data?” Internet. <http://www.ibm.com/big-data/us/en/>. Acceso: (16 de julio de 2015).
- ✓ Lantares Solutions (Sitio Web). *Big Data: ventajas de “la revolución de los datos masivos”* Internet. <http://www.lantares.com/blog/big-data-ventajas-de-la-revolucion-de-los-datos-masivos>. Acceso: (19 de julio de 2015).
- ✓ MALACARA, Nancy. *5 características del Big Data*. Internet. <http://www.informabtl.com/5-caracteristicas-del-big-data/>. Acceso: (20 de julio de 2015).
- ✓ MAYER-SCHÖNBERGER, Viktor & CUKIER, Kenneth. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Nueva York, Editorial Houghton Mifflin Hartcourt, 2ª edición, 2013.
- ✓ MEDINA LA PLATA, Edison. *Business Intelligence: Una guía práctica*. Lima, Editorial YoPublico S.A.C, 2ª edición, 2015.
- ✓ Monografías.com (Sitio Web). *Comercialización: Sistemas y Organizaciones*. Internet. <http://www.monografias.com/trabajos/comercializa/comercializa.shtml>. Acceso: (21 de julio de 2015).
- ✓ MORA PÉREZ, José. *Capacity Planning IT*. Madrid, 2ª edición, 2012.

- ✓ NAVARRETE, Óscar. TECUATL, Graciela & GONZÁLEZ, Guadalupe. *Software propietario vs software libre: una evaluación de sistemas integrales para la automatización de bibliotecas*. Internet. http://www.scielo.org.mx/scielo.php?pid=S0187-358X2011000200003&script=sci_arttext#nota. Acceso: (3 de octubre de 2015).
- ✓ NEWTON, Chris. *Diferencia entre estrategia de comercialización diferenciada y estrategia de comercialización concentrada*. Internet. <http://pyme.lavoztx.com/diferencia-entre-estrategia-de-comercializacin-diferenciada-y-estrategia-de-comercializacin-concentrada-6048.html>. Acceso: (24 de julio de 2015).
- ✓ OLVERA, Carlos. “Medidas de Almacenamiento de Información – Almacenamiento de Datos Parte II”. Internet. <https://tiposdecomputadora.wordpress.com/2010/12/04/almacenamiento-de-datos-parte-ii-medidas-de-almacenamiento-de-informacion/>. Acceso: (20 de julio de 2015).
- ✓ Pentaho. Internet. <http://www.pentaho.com/>. Acceso: (11 de octubre de 2015).
- ✓ Qlik. Internet. <http://www.qlik.com/>. Acceso: (11 de octubre de 2015).
- ✓ SABHERWAL, Rajiv & BECERRA-FERNÁNDEZ, Irma. *Business Intelligence: Practices, Technologies, and Management*. John Wiley & Sons Inc. 2ª edición, 2008.
- ✓ Silicon Week (Sitio Web). *Las cinco ventajas competitivas que aporta Big Data*. Internet. <http://www.siliconweek.es/workspace/las-cinco-ventajas-competitivas-que-aporta-el-big-data-49286>. Acceso: (21 de julio de 2015).
- ✓ Sinergia e Inteligencia de Negocio S.L. (Sitio Web). *¿Qué es Business Intelligence?* Internet. http://www.sinnexus.com/business_intelligence/. Acceso: (22 de agosto de 2015).
- ✓ SOTO, Beatriz. *¿Qué es el marketing masivo?* Internet. <http://www.gestion.org/marketing/31557/que-es-el-marketing-masivo/>. Acceso: (24 de julio de 2015).

“ANÁLISIS COMPARATIVO DE HERRAMIENTAS DE SOFTWARE LIBRE Y
PROPIETARIO PARA LA GESTIÓN DE BIG DATA EN EMPRESAS DE
COMERCIALIZACIÓN MASIVA”

- ✓ TodoBI (Sitio Web). *Comparativa QlikView vs Pentaho*. Internet.
<http://todobi.blogspot.com/2012/09/comparativa-qlikview-vs-pentaho.html>.
Acceso: (1 de octubre de 2015).
- ✓ VÁZQUEZ, Diego. *Mercados de Consumo*. Internet.
<http://www.mktdiegovazquez.com/2011/02/mercados-de-consumo.html>.
Acceso: (25 de julio de 2015).
- ✓ ZÁRATE GALLARDO, Elizabeth. *Inteligencia de Negocios*. Internet.
<http://www.gestiopolis.com/inteligencia-de-negocios/>. Acceso: (30 de agosto de 2015).

ANEXOS

Anexo 1: Glosario de siglas y términos técnicos

- **Activos:** Se denominan como el conjunto de bienes económicos que posee una organización o empresa.
- **Ad hoc:** Expresión que indica que algo es apropiado o acorde a un determinado fin.
- **APIs:** Abreviación en el idioma inglés de Application Programming Interface, traducido al español como La interfaz de programación de aplicaciones.
- **BI:** Abreviación en el idioma inglés de Business Intelligence, traducido al español como Inteligencia de Negocios.
- **Competitividad:** Se denomina como la capacidad que tiene una empresa para mantener ventajas competitivas que le permitan alcanzar una posición determinada dentro del entorno económico
- **Depuración:** Se denomina como la corrección y eliminación de posibles errores dentro de un contexto específico.
- **DIKW:** Abreviación en el idioma inglés de Data, Information, Knowledge and Wisdom, traducido al español como Datos, Información, Conocimiento y Sabiduría.
- **DWH:** Abreviación en el idioma inglés de Data Warehouse, traducido al español como almacén de datos.
- **ETL:** Abreviación en el idioma inglés de Extract, Transform and Load, traducido al español como Extracción, Transformación y Carga.
- **Explotación:** Proceso destinado a sacar el máximo provecho de algún elemento en concreto.
- **Hadoop:** Es un proyecto de código abierto utilizado exclusivamente para el almacenamiento y la distribución de grandes conjuntos de datos que soporta aplicaciones distribuidas.
- **Indicadores:** Se denomina como una medida estadística que es utilizada para indicar o señalar la relación que existe entre un conjunto de parámetros determinado.
- **KPI:** Abreviación en el idioma inglés de Key Performance Indicator, traducido al español como Indicador Clave de Desempeño.

- **Márketing:** Concepto en el idioma inglés traducido al español como Mercadotecnia. Se denomina como la disciplina que se encarga de analizar el comportamiento de los mercados y los consumidores.
- **Metadatos:** Se denominan a los datos que describen otros datos con el fin de ayudar a identificar objetos que contienen información valiosa dentro de un contexto determinado.
- **Métricas:** Se denominan como unidades de medida utilizadas para el cálculo de resultados que provienen de un previo análisis sobre una situación en concreto.
- **OLAP:** Abreviación en el idioma inglés de On-Line Analytical Processing, traducido al español como Procesamiento Analítico en Línea.
- **OS:** Abreviación en el idioma inglés de Operating System, traducido al español como Sistema Operativo.
- **PDI:** Abreviación en el idioma inglés de Pentaho Data Integration, traducido al español como Integración de Datos de Pentaho.
- **Productividad:** Se define como la relación entre la cantidad de bienes o servicios producidos en una empresa con la cantidad de recursos utilizados para producirlos.
- **PYMEs:** Abreviación de Pequeñas y Medianas Empresas.
- **Queries:** Concepto en el idioma inglés traducido al español como consultas realizadas contra una base de datos.
- **Riesgo:** Situación o escenario en el que puede ocurrir alguna desgracia o contratiempo donde generalmente se toman medidas para mitigarlo.
- **Segmentación:** Se define como el proceso de dividir o categorizar un elemento en grupos más pequeños que contengan características similares.
- **Smartphones:** Concepto en el idioma inglés, traducido al español como teléfonos inteligentes.
- **Software:** Se define como el conjunto de programas e instrucciones que permiten llevar a cabo diferentes tareas dentro de una computadora.
- **TI:** Abreviación de Tecnologías de la Información.
- **Tweets:** Publicaciones de cualquier índole propios y exclusivos de la red social Twitter.

- **Universalización:** Se define como el proceso que implica una transformación de algo en universal o general.
- **Virtuoso:** Se denomina como el contexto que gira en torno a la virtud.